Dynamical Analysis of Complex Systems

7CCMCS04

A Annibale

Department of Mathematics King’s College London

January 2011
Contents

Introduction and Overview 7
  0.1 Brownian Motion .......................................................... 7
    0.1.1 Einstein’s explanation .............................................. 7
    0.1.2 Langevin’s equation .............................................. 10
  0.2 Fokker-Planck equation ............................................... 13
  0.3 Birth-Death processes .................................................. 13

1 Probability theory 15
  1.1 Probability density and moments ..................................... 15
  1.2 Transformation of variable .......................................... 16
  1.3 Characteristic Function and Cumulants ............................. 17
  1.4 The Gaussian distribution ........................................... 18
  1.5 Generalization to several random variables ....................... 18
  1.6 The multivariate Gaussian distribution ........................... 20
  1.7 Examples and exercises ............................................. 21

2 Master equation 23
  2.1 Stochastic processes .................................................. 23
    2.1.1 Stationary processes ........................................... 25
    2.1.2 Ergodic processes .............................................. 26
  2.2 Markov processes and the Chapman-Kolmogorov equation ......... 26
  2.3 The Master equation .................................................. 27
    2.3.1 Steady state .................................................... 29
    2.3.2 Convergence to equilibrium .................................... 30
    2.3.3 Mean-field equation ........................................... 31
  2.4 Examples and exercises ............................................. 32
    2.4.1 The Wiener-Levy process ...................................... 32
    2.4.2 The Ornstein-Uhlenbeck process ............................... 32
    2.4.3 The random walk in one dimension ............................ 34

3 Markov Chains and One step processes 37
  3.1 Markov Chains ........................................................... 37
    3.1.1 Eigenvalues and eigenvectors of stochastic matrices ........ 40
    3.1.2 Definitions based on accessibility of states ............... 41
    3.1.3 Convergence to a stationary state ............................ 42
    3.1.4 Time-reversal and detailed balance ............................ 43
### CONTENTS

3.2 Examples and exercises on Markov chains ........................................ 45  
3.2.1 Two-state Markov chain ......................................................... 45  
3.2.2 Three-state chain .............................................................. 48  
3.3 One-step processes ............................................................... 52  
3.3.1 The Poisson process .......................................................... 53  
3.4 Examples and exercises on One step processes .................................. 55  
3.4.1 The decay process ............................................................. 55  
3.4.2 Birth-death process ............................................................ 55  
3.4.3 Chemical reaction .............................................................. 56  
3.4.4 Continuos time random walk .................................................. 56  
3.4.5 Diffusion-annihilation on complex networks .............................. 57  

4 Topics in equilibrium and nonequilibrium statistical mechanics ........ 59  
4.1 Equilibrium ................................................................. 60  
4.2 Probability distributions in dynamical systems .............................. 60  
4.3 Detailed balance .............................................................. 62  
4.4 Time dependent correlations ..................................................... 64  
4.5 Wiener-Khintchine theorem ...................................................... 64  
4.5.1 White noise ................................................................. 65  
4.6 Fluctuation-Dissipation theorem ............................................... 66  
4.6.1 Linear response .............................................................. 66  
4.6.2 The fluctuation-dissipation theorem ..................................... 69  
4.6.3 Power absorption ............................................................ 70  

5 Fokker Planck ................................................................. 73  
5.1 Kramers-Moyal expansion ....................................................... 73  
5.1.1 The jump moments ......................................................... 74  
5.1.2 Expression for the multivariate case ................................... 75  
5.1.3 Deterministic processes: Liouville equation ........................... 76  
5.2 Fokker-Planck as a continuity equation ..................................... 76  
5.3 Macroscopic equation .......................................................... 77  
5.4 Examples of Fokker-Planck equations ....................................... 78  

6 Langevin equation ............................................................ 81  
6.1 Langevin equation for one variable ......................................... 81  
6.2 Langevin equation for Brownian motion .................................... 82  
6.2.1 Mean-square displacement ............................................... 83  
6.2.2 Stationary velocity distribution function ............................... 84  
6.2.3 Conditional probability ..................................................... 85  
6.3 Colored noise ................................................................. 86  
6.3.1 Derivation of the Fokker-Planck equation ............................ 86  
6.4 Quasilinear Langevin equation ............................................... 88  
6.5 Non-linear Langevin equation ................................................ 89  
6.5.1 Definition of the Stochastic integral .................................. 91  
6.6 The Kramers-Moyal coefficients for the Langevin equation ............ 92  
6.6.1 Multivariate case .......................................................... 94
Introduction and Overview

A complete description of the dynamical evolution of a macroscopic system ($N$ interacting particles in a box or $N$ interacting objects in a lattice) would in principle require solving all the microscopic equations of the system. We cannot generally do this, instead we use a stochastic description, i.e. we describe the system by macroscopic variables, which fluctuate in time. Fluctuations, generally arising from system’s interactions with its environment (the “thermal bath”), the finite number of system’s discrete constituents, etc., constitute a background of “noise” which imposes limitations on the possible accuracy of physical observables measurements. These will generally give rise to data exhibiting a well defined deterministic trend, which is reproducible, with fluctuations around it, which are not.

In this course we study methods to determine the gross deterministic behaviour and the statistics of fluctuations around it, for model systems that evolve in time under the effect of a process which is too complicated to be described deterministically and therefore it can only be described probabilistically. We start giving a semi-historical outline of how a phenomenological theory of fluctuating phenomena arose and introduce some of the concepts and methods we shall be learning about.

0.1 Brownian Motion

The observation that, when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion was first investigated by English botanist Robert Brown, in 1827; as a result, the phenomenon took on the name of Brownian motion. This phenomenon reveals very clearly the statistical fluctuations which occur in a system in thermal equilibrium and can serve as a prototype problem whose analysis provides considerable insights into the mechanism responsible for the existence of fluctuations and dissipation of energy. Brownian motion is also of great practical interest because such fluctuations constitute a background of “noise” which imposes limitations on the possible accuracy of delicate physical measurements. The first convincing explanation of Brownian motion was given by Einstein in 1905.

0.1.1 Einstein’s explanation

Einstein assumed that Brownian motion is caused by frequent impacts on the pollen grains by molecules of the liquid in which it is suspended and that the motion of these molecules is so complicated that its effect on the pollen grain can only be described probabilistically in terms of frequent statistically independent impacts.

In Einstein’s model, he assumes that each individual particle executes a motion which is inde-
dependent of the motion of all other particles and that the movements of the same particle in
different time intervals are independent processes, as long as these intervals are not chosen too
small. In other words, it is assumed that impacts happen at times 0, τ, 2τ, · · · , with τ being very
small compared to the observable time intervals, but nevertheless so large that in two successive
time intervals τ the movements executed by the particle can be thought of as events which are
independent of each other.

For simplicity, consider the case of Brownian motion in one dimension, i.e. along the X-axis.
It is assumed that when an impact occurs, the probability that the pollen particle receives
a displacement between ∆ and ∆ + d∆ is

\[ \phi(\Delta) d\Delta, \]

so that

\[ \int_{-\infty}^{\infty} d\Delta \phi(\Delta) = 1. \]

It is also assumed that the probability density \( \phi \) is even, i.e.

\[ \phi(\Delta) = \phi(-\Delta), \]

from which it follows

\[ \int_{-\infty}^{\infty} d\Delta \phi(\Delta) \Delta^{2n+1} = 1 \quad \forall \quad n = 1, 2, \ldots. \]

Now suppose that \( f(x,t) dx \) is the probability that the particle lies between \( x \) and \( x + dx \) at time
\( t \), so that \( \int_{-\infty}^{\infty} dx f(x,t) = 1 \); the probability that the particle lies between \( x \) and \( x + dx \) at time
\( t + \tau \) is

\[ f(x, t + \tau) dx = \sum_{\Delta} \text{Prob}(\text{particle at } x + \Delta \text{ at time } t \text{ and receives a kick } -\Delta \text{ in the time gap } \tau) \]

\[ = dx \int_{-\infty}^{\infty} d\Delta f(x + \Delta, t)\phi(-\Delta) = dx \int_{-\infty}^{\infty} d\Delta f(x + \Delta, t)\phi(\Delta) \quad (1) \]

Since \( \tau \) is small

\[ f(x, t + \tau) = f(x, t) + \tau \frac{\partial f}{\partial t} \]

Furthermore we develop \( f(x + \Delta, t) \) in powers of \( \Delta \)

\[ f(x + \Delta, t) = f(x, t) + \Delta \frac{\partial f}{\partial \Delta} + \frac{1}{2} \Delta^2 \frac{\partial^2 f}{\partial \Delta^2} + \cdots \quad (2) \]

Using this series under the integral we get

\[ f(x, t) + \tau \frac{\partial f}{\partial \tau} = f(x, t) \int_{-\infty}^{\infty} d\Delta \phi(\Delta) + \frac{\partial f}{\partial x} \int_{-\infty}^{\infty} d\Delta \phi(\Delta)\Delta + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \int_{-\infty}^{\infty} d\Delta \phi(\Delta)\Delta^2 + \cdots \]

Terms where \( \Delta \) has an odd exponent vanish because of the parity of function \( \phi \). Out of the
remaining terms, each one is very small compared with the previous, since \( \phi \) will only be different
from zero for very small values of \( \Delta \), so we obtain

\[ \frac{\partial f}{\partial t} = D_E \frac{\partial^2 f}{\partial x^2} \quad (3) \]

where we have used normalization of \( \phi \) and set

\[ D_E = \frac{1}{2\tau} \int_{-\infty}^{\infty} d\Delta \phi(\Delta)\Delta^2 \]

This is known as the differential equation of diffusion, with \( D_E \) being the diffusion coefficient,
and its solution is, assuming the particle is at time \( t = 0 \) in \( x = 0 \),

\[ f(x, t) = \frac{1}{\sqrt{4\pi D_E t}} e^{-x^2/4D_E t} \quad (4) \]
0.1. BROWNIAN MOTION

(Note that \( f(x, t) \) is normalized, so that it is indeed a probability density).

The result (4) is obtained by Fourier transform methods: we note that equation (3) is solved by functions of the form

\[
f(x, t) = \int \frac{d\xi}{\sqrt{2\pi}} F(\xi)e^{-ix\xi - tDE\xi^2}
\]

with \( F(\xi) \) an arbitrary function of \( \xi \). Let \( f(x, 0) = h(x) \). Then

\[
F(\xi) = \int \frac{dx}{\sqrt{2\pi}} h(x)e^{ix\xi}
\]

If the particle is in \( x = 0 \) at \( t = 0 \), one has \( h(x) = \delta(x) \) and \( F(\xi) = 1/\sqrt{2\pi} \); \( f(x, t) \) follows by solving the Gaussian integral

\[
f(x, t) = \int \frac{d\xi}{2\pi} e^{-ix\xi - tDE\xi^2} = \frac{1}{\sqrt{4\pi D_E t}} e^{-x^2/4D_E t}
\]

The displacement on the X-axis that a particle experience on average leads to Einstein’s much acclaimed relation for the diffusion coefficient

\[
\lambda_x = \sqrt{\langle x^2 \rangle} = \sqrt{2D_E t}
\]

Einstein’s derivation is based on a discrete time assumption, that impacts occur at times \( 0, \tau, 2\tau, ... \), and the resulting equation must be regarded as an approximation in which \( \tau \) is considered so small that \( t \) may be considered as continuous. This description of Brownian motion contains many major concepts which are central of dynamical analysis of stochastic processes

- Chapman-Kolmogorov equation occurs in (1), where the probability of the particle being in \( x \) at time \( t + \tau \) is given by the sum of the probability of all possible “pushes” \( \Delta \) from position \( x + \Delta \), multiplied by the probability of being at \( x + \Delta \) at time \( t \). This is based on the independence of the pushes on the previous history, i.e. on the Markovian property of the process.

- Kramers-Moyal and similar expansions are essentially the approximation used in (2), which effectively replaces a process whose sample path need not be continuous with one whose paths are continuous.

- equation (3) is a special case of the Fokker-Planck equation, which is an equation for the probability distribution to find the system in a certain state, and describes a large class of very interesting stochastic processes in which the system has a continuous sample path.

The pollen’s grain position, if thought of as obeying a probabilistic law given by solving the diffusion equation, where \( t \) is continuous, can be written as \( x(t) \), where \( x(t) \) is a continuous function of time, but a random function. This leads us to consider the possibility of describing the dynamics of the system in a more direct probabilistic way, in which one has a random or stochastic differential equation for the path. This procedure was initiated by Langevin.
0.1.2 Langevin’s equation

A particle of mass $m$ and radii $a$, suspended in a fluid with viscosity $\eta$ experiences a viscous drag force in the X-direction

$$ f_s = -\alpha v = -6\pi\eta av $$

which represents the average macroscopic effect of the irregular impacts of the molecules of the fluid on the particle. The momentum of the particle is transferred to the molecules of the fluid and if the mass of the particle is large enough so that the velocity acquired by the particle because of thermal fluctuations is negligible, the dynamical equation for the particle is given by

$$ m\dot{v} = -\alpha v $$ \hspace{1cm} (5)

and the velocity of the particle will decrease to zero with the relaxation time $\tau = m/\alpha$, according to $v(t) = v(0)e^{-t/\tau}$.

If the mass is small, the thermal velocity may be observable and therefore the evolution of the particle cannot be described by (5). Nevertheless, if the mass of the particle is still large compared to that of molecules, one expects (5) to be valid approximately. However, it must be modified somehow. The modification consists in adding a fluctuating force on the RHS of (5), i.e. the total force of the molecules acting on the small particle is decomposed into a continuous damping (average) force $f_s$ and a fluctuating force $f_r(t)$ mimicking the effect of the deviation from average behaviour of molecules of the fluid. The resulting dynamical equation of the particle is

$$ m\dot{v} = -\alpha v + f_r(t) $$ \hspace{1cm} (6)

The force $f_r(t)$ is a stochastic or random force, the properties of which are given only on average.

Where does the stochastic force arise? If we were to treat Brownian motion exactly, we should solve the coupled equation of motion for all the molecules of the fluid and for the small particle. Because of the large number of molecules in the fluid (order $10^{23}$), however, we cannot generally solve these coupled equations. Furthermore, since we do not know the initial values of all the molecules of the fluid we cannot calculate the exact motion of the small particle immersed in the fluid. If we were to use another system (particle and fluid) identical to the first except for the initial values of the fluid, a different motion of the small particle would result. As usually done in thermodynamics, we consider an ensemble of such systems (Gibbs ensemble). The force $f_r(t)$, then varies from system to system and the only thing that we can do is to consider averages of this force for the ensemble.

Introducing the fluctuating force per unit mass $\Gamma(t) = f_r(t)/m$ and setting $\gamma = \alpha/m$ we get from (6)

$$ \dot{v} + \gamma v = \Gamma(t) $$ \hspace{1cm} (7)

which is a stochastic differential equation, i.e. a differential equation with a random term $\Gamma$ (the Langevin force) and hence whose solution is, in some sense, a random function. Each solution represents a different trajectory, however measurable results can be derived by using general properties of $\Gamma$. 
Since the Langevin equation contains a frictional force $-\alpha v$, it implies the existence of a process whereby the energy associated with the coordinate $x$ of the particle is dissipated in the course of time to the other degrees of freedom (e.g. to the molecules of the liquid surrounding the particle). This is in agreement with our macroscopic experience of frictional forces. Nevertheless, one deals here with an example of an interesting and conceptually difficult general problem. The \textit{microscopic equations} governing the motion of the combined system particle + fluid do not involve any frictional forces. The total energy is conserved and the motion is reversible (if the sign of the time were reversed the equation of motion would be unchanged and all particles would classically retrace their paths in time). But if one focuses attention on the massive particle, its interaction with the fluid (large system) can be adequately described by equations of motion involving frictional forces. There is thus dissipation of energy from the particle to the heat reservoir (fluid) and the motion of the particle is not reversible. It is beyond the scope of this course to understand how this situation comes about and how the modified equations of motion for the particle are derived from the microscopic equations. We shall assume the validity of Langevin’s equation as an adequate phenomenological description of Brownian motion and illustrate how it can be applied to the calculation of quantities of physical interest.

In the reminder of this section we derive the diffusion coefficient of the Brownian motion. First we rewrite (7) in terms of $x$, by using $v = \dot{x}$ and multiply it by $x$, thus obtaining

$$\frac{1}{2} \frac{d^2}{dt^2}(x^2) - \left(\frac{dx}{dt}\right)^2 = -\frac{1}{2} \gamma \frac{d}{dt}(x^2) + \Gamma x$$

We assume that the average over the ensemble of $\Gamma(t)$ is zero

$$\langle \Gamma(t) \rangle = 0$$

because the equation of motion of the average velocity should be given by (5). Normally, one also assume that the collision of different molecules of the fluid with the particle are approximately independent, so if we multiply two Langevin forces at different times the average value is zero for time differences $t' - t$ which are larger than the duration time $\tau_0$ of a collision, i.e.

$$\langle \Gamma(t) \Gamma(t') \rangle = 0 \quad \text{for} \quad |t - t'| \geq \tau_0$$

Usually, the duration time $\tau_0$ of a collision is much smaller than the relaxation time $\tau$ of the velocity of the small particle. We may therefore take the limit $\tau_0 \to 0$ as a reasonable approximation, giving

$$\langle \Gamma(t) \Gamma(t') \rangle = 2D \delta(t - t')$$

where $D$ can be determined by the energy equipartition, as we will show shortly. If the mean value of the fluctuating force $\Gamma$ always vanishes, irrespective of the value of $v$ or $x$, then $\langle x\Gamma \rangle = \langle x \rangle \langle \Gamma \rangle = 0$, because of the irregularity of $\Gamma$ and assuming \textit{independence} of $x$ and $\Gamma$, i.e. assuming that the irregularities of $x$ as a function of time do not somehow conspire to be always in the same direction as those of $\Gamma$.

If the particle is assimilated to a molecule of the liquid, in equilibrium at temperature $T$, its average kinetic energy, is given by the equipartition law (in one dimension)

$$\frac{1}{2} m \langle v^2 \rangle = \frac{1}{2} kT$$
where \( k \) is Boltzman’s constant, \( T \) is the absolute temperature. Averaging (8) over a large number of different particles and using equipartition of energy one has
\[
\frac{d^2}{dt^2} \langle x^2 \rangle + \gamma \frac{d}{dt} \langle x^2 \rangle = \frac{2kT}{m}
\]
(10)
where the term \( \langle x \Gamma \rangle \) has been set equal to zero. The solution is given by (\( C \) is an integration constant)
\[
\frac{d\langle x^2 \rangle}{dt} = \frac{2kT}{\gamma m} + Ce^{-\gamma t}
\]
(11)
Langevin estimated that the decaying exponential approaches zero with a time constant of the order of \( 10^{-8} \), so that \( d\langle x^2 \rangle/dt \) enters rapidly a constant regime \( d\langle x^2 \rangle/dt = 2kT/\gamma m \). Therefore one further integration in this asymptotic regime yields
\[
\langle x^2 \rangle - \langle x^2_0 \rangle = \frac{2kT}{\gamma m} t
\]
(12)
and the particle behaves like a diffusing particle executing a random walk, so \( \langle x^2 \rangle \propto t \). Indeed equation (12) corresponds to Einstein’s relation, once we identify
\[
D_E = \frac{kT}{\gamma m}
\]
Einstein’s condition of independence of displacements \( \Delta \) at different times is equivalent to Langevin assumption \( \langle \Gamma x \rangle = 0 \).

Note that Einstein relation for the diffusion coefficient
\[
D_E = \lim_{t \to \infty} \frac{\langle x^2 \rangle - \langle x^2_0 \rangle}{2t} = \frac{kT}{\gamma m} = \frac{RT}{6N_A \pi \eta a}
\]
where \( R \) is the perfect gas constant and \( N_A \) is the Avogadro number.

allows the determination of the Avogadro number (a microscopic quantity) from experimentally accessible macroscopic quantities, thus providing a nonambiguous link between the microscopic and macroscopic levels of description in physics and can be considered as a “proof” of the existence of atoms.

The picture of a Brownian particle immersed in a fluid is typical of a variety of problems, even when there are no particles. For instance, it is the case if there is only one slowly or heavy degree of freedom that interacts, in a more or less irregular random fashion, with many other fast or light degrees of freedom, which play the role of the bath. Thus the general concept of fluctuations describable by Langevin equation and Fokker-Planck equations has developed very extensively in a very wide range of situations.

The advantages of a continuous description turn out to be very significant, since only a very few parameters are required, i.e. essentially the coefficient of the derivatives
\[
\int_{-\infty}^{\infty} d\Delta \phi(\Delta) \Delta, \quad \int_{-\infty}^{\infty} d\Delta \phi(\Delta) \Delta^2
\]
(13)
It is rare a problem (mechanical oscillators, fluctuations in electrical circuits, chemical reactions, dynamics of dipoles and spins, escape over metastable barriers, etc.) which cannot be specified,
in at least some degrees of freedom, by the corresponding Fokker-Planck equation (allowing both coefficients (13) to depend on \( x \) and \( t \), and in a space of an appropriate number of dimensions), or equivalently, by augmenting a deterministic differential equation with some fluctuating force or field, like in Langevin’s approach.

0.2 Fokker-Planck equation

The general Fokker-Planck equation for one variable \( x \) has the form

\[
\frac{\partial f(x,t)}{\partial t} = \left[ -\frac{d}{dx} D^{(1)}(x) + \frac{d^2}{dx^2} D^{(2)}(x) \right] f(x,t).
\]

\( D^{(2)}(x) > 0 \) is called the diffusion coefficient and \( D^{(1)}(x) \) is the drift coefficient. The diffusion and drift coefficient may also depend on time. Mathematically, it is a linear second-order partial differential equation of a parabolic type.

A generalization to the \( N \) variables \( x_1 \cdots x_N \) has the form

\[
\frac{d}{dt} f(\{x\}, t) = \left[ -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} D_i^{(1)}(\{x\}) + \sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\{x\}) \right] f(\{x\}, t). \tag{14}
\]

The Fokker-Planck equation is an equation for the distribution function of \( N \) macroscopic variables \( \{x\} \).

For a deterministic treatment we neglect the fluctuations of the macroscopic variables. For the FP equation this means that we neglect the diffusion term. Equation (14) is then equivalent to the system of differential equations

\[
\frac{d}{dt} x_i = D_i^{(1)}(x_1, \cdots, x_N), \quad i = 1, \cdots, N
\]

for the \( N \) macrovariables \( \{x\} \).

0.3 Birth-Death processes

A wide number of phenomena can be modelled by birth-death processes. An entertaining model is the predator-prey system (Lotka-Volterra, 1925/26), consisting of two species, one of which preys on the other, which is itself supplied with an inexhaustible food supply. Let \( X \) symbolize prey, \( Y \) the predator and \( A \) the food of the prey. The process under consideration may be

\[
X + A \rightarrow 2X \\
Y + X \rightarrow 2Y \\
Y \rightarrow B \quad \text{(death by natural causes)} \tag{15}
\]

It is easy to guess model differential equations for \( x \) and \( y \), the numbers of \( X \) and \( Y \)

\[
\frac{d}{dt} x = k_1 ax - k_2 xy \\
\frac{d}{dt} y = k_2 xy - k_3 y \tag{16}
\]
Fixed points are \( \mathbf{a}_1 = (0,0)^t \) and \( \mathbf{a}_2 = (ak_1/k_2,k_3/k_2)^t \). \( \mathbf{a}_1 \) is unstable, \( \mathbf{a}_2 \) is marginally stable. Leads to oscillatory behaviour in predator and prey population, which could be exploited economically (e.g. hare–lynx \( \Rightarrow \) price of lynx-fur high when number of lynx is low!). Of course, realistic systems do not follow the solutions of the differential equations exactly, but fluctuate about these curves. One must include these fluctuations and the easiest way to do this, is by means of a birth-death master equation. We assume a probability distribution \( P(x,y,t) \) for the number of individuals at time \( t \) and that in an infinitesimal time interval \( \Delta t \) the following transition probabilities laws hold

\[
\begin{align*}
\text{Prob}(x \to x+1, y \to y) &= k_1 ax \Delta t \\
\text{Prob}(x \to x - 1, y \to y + 1) &= k_2 xy \Delta t \\
\text{Prob}(x \to x, y \to y - 1) &= k_3 y \Delta t \\
\text{Prob}(x \to x, y \to y) &= 1 - (k_1 ax + k_2 xy + k_3 y) \Delta t
\end{align*}
\]  

(17)

So far we have simply replaced the rate laws by probability laws. Next, we employ Chapman-Kolmogorov equation, i.e. we write the probability at time \( t + \Delta t \) as the sum of the probabilities of all previous states multiplied by the probability of a transition to the state \( (x,y) \). By letting \( \Delta t \to 0 \)

\[
\frac{P(x,y,t + \Delta t) - P(x,y,t)}{\Delta t} \to \frac{\partial P(x,y,t)}{\partial t} = k_1 a(x-1)P(x-1,y,t) + k_2(x+1)(y-1)P(x+1,y-1,t) + k_3(y+1)P(x,y-1,t) - (k_1 ax + k_2 xy + k_3 y)P(x,y,t)
\]  

(18)

This model has a wide application (systems of molecules, photons, biological systems, etc.) The choice of transition probabilities can be derived with much greater detail than the simple multiplicative choice (this is just the most elementary). Solutions of (18) determine both the gross deterministic motion and the fluctuations. The latter are typically of magnitude of the square root of the number of individuals involved. A theory can be developed which can deal with a wide range of models in this category and there is a close connection between this kind of theory and that of stochastic differential equations.
Chapter 1

Probability theory

1.1 Probability density and moments

We call $X$ a stochastic variable if the number $X$ cannot be predicted. By repeating the experiment $N$ times ($N$ realizations) we obtain $N$ numbers

$$X_1, X_2, ..., X_N$$

Instead of repeating the experiment with one system $N$ times, one may also think to have an ensemble of $N$ identical systems and make one experiment for each system. Whereas the numbers $X_1, ..., X_N$ cannot be predicted, some averages for $N \to \infty$ may be predicted and should give the same value for identical systems. The simplest average is

$$\langle X \rangle = \lim_{N \to \infty} \frac{1}{N}(X_1 + ... + X_N);$$

(1.2)

A general average is

$$\langle f(X) \rangle = \lim_{N \to \infty} \frac{1}{N}(f(X_1) + ... + f(X_N))$$

(1.3)

where $f$ is some arbitrary function.

If $f(X) = \theta(x - X)$, where $\theta$ is the step function defined as

$$\theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(1.4)

and if $M$ is the number of experiments (or realizations) where the random variable $X$ is equal or less than $x$ then

$$\langle \theta(x - X) \rangle = \lim_{N \to \infty} \frac{1}{N}(\theta(x - X_1) + ... + \theta(x - X_N))$$

$$= \lim_{N \to \infty} \frac{M}{N} \equiv p(X \leq x)$$

(1.5)

i.e. in the limit $N \to \infty$ the relative frequency $M/N$ is the probability that the random variable $X$ is equal or less than $x$. 
The probability density function $P_X(x)$ of the random variable $X$ is defined, assuming that $p$ is differentiable, as
\[ P_X(x) = \frac{d}{dx} p(X \leq x) = \frac{d}{dx} (\theta(x-X)) = \langle \delta(x - X) \rangle \] (1.6)
where we introduced the Dirac $\delta$ as the derivative of the step function. Hence, the probability of finding the stochastic variable $X$ in the interval $x \leq X \leq x + dx$ is
\[ p(X \leq x + dx) - p(X \leq x) = dx \frac{d}{dx} p(X \leq x) = P_X(x) dx \] (1.7)
Normalization requires $\int dx P_X(x) = 1$.

For discrete random variables, $p$ jumps at the discrete values $x_n$ and $P_X(x)$ then consists in a sum of $\delta$ functions
\[ P_X(x) = \sum_n p_n \delta(x - x_n) \] (1.8)
where $p_n$ is the probability to find the discrete value $x_n$.

The statistical properties of the random variable $X$ are completely determined by the probability density $P_X(x)$. It is easy to form averages on the basis of the given probability density $P_X(x)$
\[ \langle f(X) \rangle = \langle \int dx \delta(x - X) f(x) \rangle = \int dx P_X(x) f(x) \] (1.9)
In particular, $\langle X^m \rangle \equiv \mu_m$ is called the $m$th moment of $X$, and $\mu_1$ the average or mean. Also,
\[ \sigma^2 = \langle (X - \langle X \rangle)^2 \rangle = \mu_2 - \mu_1^2 \] (1.10)
is called the variance, which is the square of the standard deviation $\sigma$.

Exercise - Show that the Cauchy distribution
\[ P(x) = \frac{\gamma}{\pi((x-a)^2 + \gamma^2)} \quad (-\infty < x < \infty) \] (1.11)
does not have finite variance.

1.2 Transformation of variable

Let the continuous variable $X$ be mapped into the new variable $Y = f(X)$. The probability that $Y$ has a value between $y$ and $y + dy$ is
\[ P_Y(y) dy = \int_{y < f(x) < y + dy} dx P_X(x) \] (1.12)
An equivalent formula is
\[ P_Y(y) = \int dx P_X(x) \delta[f(x) - y] \] (1.13)
The last integral is easily evaluated: if $x_n = f^{-1}_n(y)$ is the $n$th simple root of $f(x) - y = 0$, then
\[ P_Y(y) = \sum_n \int dx P_X(x) \delta(x - x_n) \frac{df}{dx}^{-1}_{x=x_n} = \sum_n P_X(x_n) |f'(x_n)|^{-1} \] (1.14)
A familiar example is the derivation, from the one-dimensional Maxwell distribution for $X = v$,

$$P(v) = \sqrt{\frac{m}{2\pi kT}} \exp\left(-\frac{mv^2}{2kT}\right)$$

of the Boltzmann probability density for the energy $E = \frac{mv^2}{2}$

$$P(E) = \frac{1}{\sqrt{\pi kT}} \exp\left(-\frac{E}{kT}\right)$$

Exercise Show the derivation of the above result.

1.3 Characteristic Function and Cumulants

The characteristic function of a stochastic variable $X$ is defined by the average of $\exp(ikX)$, namely

$$G_X(k) = \langle e^{ikX} \rangle = \int dx P(x) e^{ikX}$$

(1.17)

This is merely the Fourier transform of $P_X(x)$, and can naturally be solved for the probability distribution

$$P_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp(-ikx)G_X(k).$$

(1.18)

The function $G_X(k)$ provides an alternative complete characterization of the probability distribution. It exists for all real $k$ and has the properties

$$G(0) = 1; \quad |G(k)| \leq 1$$

(1.19)

By expanding the exponential in the integrand of (1.17) and interchanging the order of the resulting series and the integral, one gets

$$G(k) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \int dx x^m P_X(x) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \mu_m$$

(1.20)

Therefore, one finds that $G_X(k)$ is the moment generating function, in the sense that

$$\mu_m = (-i)^m \frac{\partial^m}{\partial k^m} G_X(k)|_{k=0}$$

(1.21)

The same function $G_X(k)$ also serves to generate the cumulants $\kappa_m = \langle (X^m) \rangle$ which are defined by

$$\log G(k) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} \kappa_m$$

(1.22)

Note that, owing to $P_X(x)$ is normalised, $\log G(0) = 0$ so the $m = 0$ term vanishes and the above series begins at $m = 1$. The cumulants are combinations of the moments. For the first four cumulants, for example, one has

$$\kappa_1 = \mu_1$$
$$\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3$$
$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$$

(1.23)
We finally mention that there exists a general expression for $\kappa_m$ in terms of the determinant of a $m \times m$ matrix constructed with the moments $\{\mu_i | i = 1, ..., m\}$

$$\kappa_m = (-1)^{m-1} \det \begin{pmatrix}
\mu_1 & 1 & 0 & 0 & 0 & \cdots \\
\mu_2 & \mu_1 & 1 & 0 & 0 & \cdots \\
\mu_3 & \mu_2 & 2 & \mu_1 & 1 & \cdots \\
\mu_4 & \mu_3 & 3 & \mu_2 & 3 & \mu_1 & \cdots \\
\mu_5 & \mu_4 & 4 & \mu_3 & 4 & \mu_2 & \mu_1 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\end{pmatrix}$$

(1.24)

1.4 The Gaussian distribution

The general form of the Gaussian distribution for one variable is

$$P(x) = Ce^{-\frac{1}{2}Ax^2 - Bx}$$

(1.25)

where $A$ is a positive constant that determines the width, $B$ determines the position of the peak and $C$ is a normalization constant

$$C = \left(\frac{A}{2\pi}\right)^{\frac{1}{2}} e^{-B^2/2A}$$

(1.26)

It is often convenient to express the parameters $A, B$ in term of the average $\mu_1 = -B/A$ and variance $\sigma^2 = 1/A$

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left[ -\frac{(x - \mu_1)^2}{2\sigma^2} \right]$$

(1.27)

The characteristic function is

$$G(k) = \langle e^{ikx} \rangle = \frac{1}{\sqrt{2\pi}\sigma^2} \int dx \exp \left[ -\frac{(x - \mu_1)^2}{2\sigma^2} + ikx \right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma^2} \int dx \exp \left[ -\frac{x^2}{2\sigma^2} - 2(\mu_1 + ik\sigma^2)x + \mu_1^2 \right] = e^{i\mu_1 k - \frac{1}{2}\sigma^2 k^2}$$

(1.28)

The use of cumulants is particularly suited for this distribution since

$$\kappa_1 = \mu_1, \quad \kappa_2 = \sigma^2, \quad \kappa_3 = \kappa_4 = ... = 0$$

(1.29)

1.5 Generalization to several random variables

Generally we may have $n$ random variables $X_1, ..., X_n$. Similarly to (1.6) we may introduce the $n$-dimensional distribution function

$$P_{X_1, ..., X_n}(x_1, ..., x_n) \equiv P_n(x_1, ..., x_n) = \langle \delta(x_1 - X_1)...\delta(x_n - X_n) \rangle$$

(1.30)
Averages can be calculated as
\[
\langle f(X_1, ..., X_n) \rangle = \int dx_1 \cdots \int dx_n f(x_1, ..., x_n) P_n(x_1, ..., x_n)
\] (1.31)

The probability density of \( s < n \) random variables (marginal distribution) is obtained by integration over the other variables
\[
P_s(x_1, ..., x_s) = \int dx_{s+1} \cdots \int dx_n P_n(x_1, ..., x_s, x_{s+1}, ..., x_n)
\] (1.32)

Alternatively, one may consider those realizations of the \( n \) random variables \( X_1, ..., X_n \) where the first \( s \) random variables take the fixed values \( x_1, ..., x_s \), and consider the joint probability of the remaining variables \( X_{s+1}, ..., X_n \). This is called the *conditional probability* and it is denoted by
\[
P_{n-s|s}(x_{s+1}, ..., x_n|x_1, ..., x_s)
\] (1.33)

By Bayes’ rule,
\[
P_n(x_1, ..., x_n) = P_{n-s|s}(x_{s+1}, ..., x_n|x_1, ..., x_s) P_s(x_1, ..., x_s)
\] (1.34)

and we may express the conditional probability density in terms of \( P_n \) only
\[
P_{n-s|s}(x_{s+1}, ..., x_n|x_1, ..., x_s) = \frac{P_n(x_1, ..., x_n)}{\int dx_{s+1} \cdots \int dx_n P_n(x_1, ..., x_n)}
\] (1.35)

Note that from normalization of the joint and marginal probabilities it follows the normalization of the conditional probability.

For multivariate probability distributions, the moments are defined by
\[
\langle X_1^{m_1} \cdots X_n^{m_n} \rangle = \int dx_1 \cdots dx_n x_1^{m_1} \cdots x_n^{m_n} P_n(x_1, ..., x_n)
\] (1.36)

while the characteristic (moment generating) function depends on \( n \) auxiliary variables \( k = (k_1, ..., k_n) \)
\[
G(k) = \langle \exp[i(k_1 X_1 + \cdots + k_n X_n)] \rangle = \sum \frac{(i k_1)^{m_1} \cdots (i k_n)^{m_n}}{m_1! \cdots m_n!} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle
\] (1.37)

Similarly, the cumulants of the multivariate distribution are defined in terms of the coefficients of the expansion of \( \ln G \) as
\[
\ln G(k) = \sum \frac{(i k_1)^{m_1} \cdots (i k_n)^{m_n}}{m_1! \cdots m_n!} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle
\] (1.38)

where the prime indicates the absence of the term with all the \( m_i \) simultaneously vanishing (by the normalization of \( P_n \)).
1.6 The multivariate Gaussian distribution

A collection of \( n \) random variables \( X_1, \ldots, X_n \) are said to have a zero-mean (joint) Gaussian probability distribution if

\[
P_n(x_1, \ldots, x_n) = C \exp \left( -\frac{1}{2} \sum_{ij} A_{ij} x_i x_j \right)
\]

with positive-definite symmetric matrix \( A \) and \( C \) determined from normalization as

\[
C = \sqrt{\frac{\det A}{(2\pi)^n}}
\]

In vector notation

\[
P(x) = \sqrt{\frac{\det A}{(2\pi)^n}} e^{-\frac{1}{2} x A^{-1} x}
\]

The symmetry of the distribution implies \( \langle x_i \rangle = 0 \) and the covariances are

\[
\langle x_i x_j \rangle = (A^{-1})_{ij}
\]

It follows that a zero-mean Gaussian distribution is fully determined by its covariances. Writing the covariance matrix as \( C = A^{-1} \) one can represent the distribution in the form

\[
P(x) = \frac{1}{(2\pi)^{n/2} \det C} e^{-\frac{1}{2} x C^{-1} x}
\]

One can immediately work out the characteristic function

\[
G(k) = \langle e^{ikx} \rangle = e^{-\frac{1}{2} k C k}
\]

Thus the characteristic function of a Gaussian distribution is again Gaussian.

These results generalize easily to the case of nonzero means by setting \( y = x + \mu \). The probability distribution of \( y \) is then

\[
P(y) = \frac{1}{(2\pi)^{n/2} \det C} e^{-\frac{1}{2} (y - \mu) C^{-1} (y - \mu)}
\]

with \( \langle y_i \rangle = \mu_i \) and \( \langle y_i y_j \rangle = \mu_i \mu_j + C_{ij} \), i.e. \( C_{ij} = \langle (y_i y_j) \rangle \). The corresponding characteristic function is

\[
G(k) = \langle e^{iky} \rangle = e^{ik\mu} e^{ik\mu - \frac{1}{2} k C k}
\]

and hence again of a Gaussian form.

If follows that a Gaussian distribution is fully determined by the averages of the variables and their covariance matrix. If the variables are uncorrelated then \( A^{-1} \) is diagonal, so \( A \) is diagonal and the variables are also independent (i.e. the joint probability distribution factorizes). This independence can always be achieved by a transformation of variables.
The moments of a multivariate Gaussian distribution with zero mean have a remarkable property. Consider (1.41). We rewrite its characteristic function as

\[ G(k) = \prod_{pq} e^{\frac{1}{2}(ik_q)C_{qp}(ik_p)} = \prod_{pq} e^{\frac{1}{2}(ik_p)(ik_q)(x_p x_q)} = \prod_{pq} \left[ 1 + \frac{1}{2}(ik_p)(ik_q)(x_p x_q) + \ldots \right] \quad (1.47) \]

To find the moments \( \langle x_i x_j x_k \ldots \rangle \) one has to collect the terms proportional to \( k_i k_j k_k \ldots \). We note that only terms with an even number of factors \( k \) show up in (1.47). As a consequence, all odd moments of a zero-mean Gaussian distribution vanish. Hence, we restrict to the case of an even number of factors in \( \langle \ldots \rangle \). We suppose that the subscripts of the factors are different from each other, so that the dots in (1.47) cannot contribute. Hence, the only way in which \( k_i k_j k_k \ldots \) occurs is as the result of multiplying suitable pairs of \( k_p k_q \). On the other hand, each product of pairs \( k_p k_q \) that makes up \( k_i k_j k_k \ldots \) does occur. The result is

\[ \langle x_i x_j x_k \ldots \rangle = \sum \langle x_p x_q \rangle \langle x_u x_v \rangle \ldots \quad (1.48) \]

The subscripts \( p, u, u, v \ldots \) are the same as \( i, j, k, \ldots \) taken two by two. The summation extends over all different ways in which \( i, j, k, \ldots \) can be subdivided in pairs. The factor \( 1/2 \) in (1.47) cancels because the product contains each pair twice.

For instance,

\[ \langle x_1 x_2 x_3 x_4 \rangle = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle \quad (1.49) \]

This is a realization of Wick theorem in classical systems.

### 1.7 Examples and exercises

To illustrate the definitions given for multivariate distributions, let us compute them for a simple two-variable Gaussian distribution

\[ P_2(x_1, x_2) = \sqrt{\frac{1 - \lambda^2}{(2\pi \sigma^2)^2}} \exp \left[ -\frac{1}{2\sigma^2} (x_1^2 - 2\lambda x_1 x_2 + x_2^2) \right] \quad (1.50) \]

where the parameter \(-1 < \lambda < 1\) is such to ensure that the quadratic form in the exponent is positive definite. To fix ideas, one may interpret \( P_2(x_1, x_2) \) as the Boltzman distribution of two harmonic oscillators coupled by a potential term \( \propto \lambda x_1 x_2 \).

(a) Verify that the normalization factor takes this value by direct integration or by comparing our distribution with the multidimensional Gaussian distribution (1.41) where

\[ A = \frac{1}{\sigma^2} \begin{pmatrix} 1 & -\lambda \\ -\lambda & 1 \end{pmatrix} \quad (1.51) \]

(b) Verify that the marginal probability of the individual variables (for one of them since they are identical), defined by \( P_1(x_1) = \int dx_2 \ P_2(x_1, x_2) \) is

\[ P_1(x_1) = \frac{1}{\sqrt{2\pi \sigma_\lambda^2}} \exp \left( -\frac{x_1^2}{2\sigma_\lambda^2} \right) \quad \text{with} \quad \sigma_\lambda^2 = \frac{\sigma^2}{1 - \lambda^2} \quad (1.52) \]
CHAPTER 1. PROBABILITY THEORY

We see that the marginal distribution depends on $\lambda$, which results in a modified variance. To see that $\sigma_1^2$ is indeed the variance $\langle x_1^2 \rangle = \langle x_1^2 \rangle - \langle x_1 \rangle^2$, note that $\langle x_1^m \rangle$ can be obtained from the marginal distribution only (this is a general result)

$$\langle x_1^m \rangle = \int dx_1 dx_2 x_1^m p_2(x_1, x_2) = \int dx_1 x_1^m p_1(x_1)$$ (1.53)

Then, inspecting the marginal distribution, verify that

$$\langle x_1 \rangle = 0 \quad \langle x_2 \rangle = 0$$ (1.54)

$$\langle x_1^2 \rangle = \sigma_1^2 \quad \langle x_2^2 \rangle = \sigma_2^2$$ (1.55)

(c) To complete the calculation of the moments up to second order, we need the covariance of $x_1$ and $x_2$: $\langle x_1 x_2 \rangle = \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle$ which reduces to calculate $\langle x_1 x_2 \rangle$. Show that

$$\langle x_1 x_2 \rangle = \frac{\lambda}{1 - \lambda^2} \sigma^2$$ (1.56)

(d) It is convenient to calculate the normalised variance, or correlation coefficient,

$$\gamma = \frac{\langle x_1 x_2 \rangle}{\sqrt{\langle x_1^2 \rangle \langle x_2^2 \rangle}}$$ (1.57)

which is merely given by $\lambda$. Therefore the parameter $\lambda$ in the distribution is a measure of how correlated the variables $x_1$ and $x_2$ are. Note that in the limit $\lambda \to 0$ the variables are not correlated at all and the distribution factorizes

$$p_2|_{\lambda=0} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_1^2/2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_2^2/2\sigma^2}$$ (1.58)

In the limit $\lambda \to 1$ the variables are maximally correlated, $\gamma = 1$. The distribution becomes a function of $x_1 - x_2$, so it is favoured that $x_1$ and $x_2$ take similar values

$$p_2|_{\lambda=1} \to e^{-(x_1-x_2)^2/2\sigma^2}$$ (1.59)

We can now interpret the increase of the variance with $\lambda$: the correlation between the variables allow them to take arbitrarily large values, with the only restriction of their difference being small.

(e) By using Bayes rule $p_{1|1}(x_1|x_2) = p_2(x_1, x_2)/p_1(x_2)$ show that

$$p_{1|1}(x_1|x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x_1 - \lambda x_2)^2 \right]$$ (1.60)

Then, at $\lambda = 0$ (no correlation) the values taken by $x_1$ are independent of $x_2$, while for $\lambda \to 1$ they are centered around those taken by $x_2$, and hence strongly conditioned by them.
Chapter 2

Master equation

Now that we have reviewed some of the basic ideas of probability theory, we will begin to study how probability distributions evolve in time. In this chapter we will derive equations for the evolution of probability distribution for processes in which most of the memory effects in the evolution can be neglected (Markov processes) and we shall apply these equations to a variety of problems involving discrete stochastic variables. The level at which we discuss the evolution of probability distributions in this chapter is semiphenomenological because we do not concern ourselves with the underlying dynamics of the stochastic variables. We approximate the dynamics by a judicious choice of the transition matrix.

The equations which govern the evolution of the probability distribution for Markov processes is the Master equation. It is one of the most important equations in statistical physics because of its almost universal applicability. As a system of stochastic variables evolves, there are transitions between various values of the stochastic variables. Through transitions, the probability of finding the system in a given state changes until the system reaches a final equilibrium steady state in which transitions cannot cause further changes in the probability distributions (it can happen that it never reaches a steady state, but we will be concerned mostly with the cases for which it can). To derive the Master equation we must assume that the probability for each transition depends only on the preceding step and not on any previous history (this assumption applies to many systems). If the transitions between values of the stochastic variable only occur in small steps, then the master equation reduces approximately to a partial differential equation for the probability density (one example for such a partial differential equation is the Fokker-Planck equation that we will derive in Chapter 5).

2.1 Stochastic processes

A stochastic process is the time evolution of a stochastic variable. If $X$ is the stochastic variable, the stochastic process is $X(t)$. A stochastic variable is defined by specifying the set of possible values called “range” or “set of states” and the probability distribution over the set. The set can be discrete (eg. number of electrons arrived at an anode up to a certain time) or continuous (eg. the velocity of a Brownian particle). If the set is multidimensional the stochastic variable is a vector (eg. the three velocity components of a Brownian particle).

Stochastic processes are often encountered in Statistical Mechanics. The latter studies systems
CHAPTER 2. MASTER EQUATION

of large numbers of particles, so precise calculations cannot be made. Instead, average values are measured and fluctuations about the average values are calculated by means of probability theory. An archetypal example of stochastic process is the Brownian motion, i.e. the motion of a heavy colloidal particle immersed in a fluid made up of light particles. The stochastic variable \( X \) in this case may be the position or the velocity of the particle.

A standard procedure in thermodynamics is to regard the system as an instance of an ensemble of systems, i.e. a collection of systems each leading to a number of large numbers of particles, so precise calculations cannot be made. Instead, average values are measured and fluctuations about the average values are calculated by means of probability theory. An archetypal example of stochastic process is the Brownian motion, i.e. the motion of a heavy colloidal particle immersed in a fluid made up of light particles. The stochastic variable \( X \) in this case may be the position or the velocity of the particle.

A standard procedure in thermodynamics is to regard the system as an instance of an ensemble of systems, i.e. a collection of systems each leading to a number \( X \) which depends on time. Although the outcome of one system cannot be precisely predicted, we assume that ensemble averages exist and can be calculated. For the fixed time \( t_1 \) we may therefore define a probability density by

\[
P_1(x_1, t_1) = \langle \delta(x_1 - X(t_1)) \rangle
\]

where the bracket \( (...) \) indicates the ensemble average. Similarly we can define the probability density \( P_n \)

\[
P_n(x_1, t_1; ..., x_n, t_n) = \langle \delta(x_1 - X(t_1))...\delta(x_n - X(t_n)) \rangle
\]

If we know the infinite hierarchy of probability densities

\[
P_1(x_1, t_1), \ P_2(x_1, t_1; x_2, t_2), \ P_3(x_1, t_1; x_2; t_2; x_3, t_3), \ ...
\]

we know completely the time dependence of the process described by the random variable \( X(t) \).

The hierarchy of functions \( P_n \) obey the following consistency conditions:

1. \( P_n \geq 0; \)
2. \( P_n \) does not change on interchange of two pairs \( (x_k, t_k) \) and \( (x_\ell, t_\ell) \)
3. \( \int dx_n P_n(x_1, t_1; ..., x_{n-1}, t_{n-1}; x_n, t_n) = P_{n-1}(x_1, t_1; ..., x_{n-1}, t_{n-1}) \)
4. \( \int dx_1 P_1(x_1, t_1) = 1 \)

We can obtain any average from the probability density, as in Sec. 1.5. The only new feature is that these average values are no longer just numbers, but are now functions of the \( t \) values specifying which random variables are being averaged. For example, the average of \( X \) at time \( t_1 \) is given by

\[
\langle X(t_1) \rangle = \int dx_1 P_1(x_1, t_1)x_1
\]

More generally, take \( n \) values \( t_1, t_2, ..., t_n \) for the time variable (not necessarily different) and form the \( n \)-th moment

\[
\langle X(t_1)X(t_2)...X(t_n) \rangle = \int dx_1dx_2...dx_n P_n(x_n, t_n; ..., x_2, t_2; x_1, t_1)x_1x_2...x_n
\]

Of particular interest is the autocorrelation function

\[
\kappa(t_1, t_2) = \langle \langle X(t_1)X(t_2) \rangle \rangle = \langle X(t_1)X(t_2) \rangle - \langle X(t_1) \rangle \langle X(t_2) \rangle
\]

For \( t_1 = t_2 \) it reduces to the time-dependent variance \( \langle X^2(t) \rangle \). The prefix auto is used because \( X(t_1), X(t_2), ... \) belong to the same random process. If two random processes, say \( X(t) \) and \( Y(t) \), are involved in some problem, then expectations such as \( \langle X(t_1)Y(t_2) \rangle \) may arise. The resulting function of \( t_1 \) and \( t_2 \) is called correlation function.
2.1. STOCHASTIC PROCESSES

The generalization of the concept of a characteristic function to stochastic processes is the characteristic functional. Introduce an arbitrary auxiliary test function \( k(t) \), then the characteristic or moment generating functional is defined as the following functional of \( k(t) \)

\[
G([k]) = \langle \exp \left[ i \int_{-\infty}^{\infty} dt\, k(t)X(t) \right] \rangle
\]

(2.7)

The notation \( G([k]) \) emphasizes that \( G \) depends on the whole function \( k \). Expand in powers of \( k \) and get

\[
G([k]) = \sum_{m=0}^{\infty} \frac{i^m}{m!} \int dt_1...dt_m k(t_1)...k(t_m)\langle X(t_1)...X(t_m) \rangle
\]

(2.8)

Thus each moment of the joint distribution of \( X(t_1), X(t_2), ... \) can be found as the coefficient of the term with \( k(t_1)k(t_2)... \) in this expression. Similarly, the cumulants can be found from

\[
\log G([k]) = \sum_{m=1}^{\infty} \frac{i^m}{m!} \int dt_1...dt_m k(t_1)...k(t_m)\langle X(t_1)...X(t_m) \rangle
\]

(2.9)

As done in Sec.1.5, we may attribute fixed values \( X_1, ..., X_s \) to the random variable \( X \) at the times \( t_1, ..., t_s \) and consider the joint probability for the variable \( X \) to assume the values \( X_{s+1}, ..., X_n \) at the later times \( t_{s+1}, ..., t_n \), under the condition that at the times \( t_1, ..., t_s \) it had the sharp values \( x_1, ..., x_s \),

\[
P(x_{s+1}, t_{s+1}; ..., x_n, t_n|x_s, t_s; ...; x_1, t_1) = \langle \delta(x_n - X(t_n))...\delta(x_{s+1} - X(t_{s+1})) \rangle_{X_{t_1}=x_1, ..., X_{t_s}=x_s}
\]

with \( t_n > t_{n-1} > ... > t_1 \)

(2.10)

Again

\[
P_{n-s} (x_{s+1}, t_{s+1}; ..., x_n, t_n|x_s, t_s; ...; x_1, t_1) = \frac{P_n(x_n, t_n; ...; x_1, t_1)}{\int dx_1... \int dx_s P_n(x_n, t_n; ...; x_1, t_1)}
\]

(2.11)

2.1.1 Stationary processes

A process is stationary if all \( P_n \) are not changed by replacing \( t_i \) with \( t_i + T \) (\( T \) arbitrary)

\[
P_n(x_1, t_1; x_2, t_2; ...; x_n, t_n) = P_n(x_1, t_1 + T; x_2, t_2 + T; ...; x_n, t_n + T)
\]

(2.12)

It then follows that \( P_1 \) cannot depend on \( t \) and that all the joint probability distributions \( P_n, n \geq 2 \), can only depend on the time differences \( t_k - t_{k-1} \) for all \( ts \) which occur in them. As a consequence, the mean value of a stationary stochastic process is independent of time (i.e. \( \langle X(t) \rangle \) is constant) and all moments depend only on time differences and are not affected by a shift in time

\[
\langle X(t_1)X(t_2)...X(t_n) \rangle = \langle X(t_1 + T)X(t_2 + T)...X(t_n + T) \rangle
\]

(2.13)

Note that the time independence of \( P_1 \) (or of the average \( \langle X \rangle \)) is a necessary but by no means sufficient condition for the stationarity of the process. All physical processes in equilibrium are stationary. For zero-average stationary processes the correlation function is

\[
\kappa(\tau) = \langle X(t)X(t + \tau) \rangle = \langle X(0)X(\tau) \rangle = \langle X(-\tau)X(0) \rangle = \kappa(-\tau)
\]

(2.14)

Often there exists a constant \( \tau_C \) such that \( \kappa(\tau) \simeq 0 \) for \( \tau > \tau_C \). One calls \( \tau_C \) the autocorrelation time of the stationary stochastic process.
2.1.2 Ergodic processes

When dealing with stochastic processes, what is usually available at one time is the time average of a given measurable quantity \( X(t) \) which fluctuate with time. However, what we have considered above are ensemble averages, in which we repeat the same measurement for many different copies of our system and compute averages \( \langle \ldots \rangle \). For very many processes, i.e. ergodic processes, the ensemble average is equivalent to the time average over an arbitrary large time \( T \), which is then allowed to become infinite.

To arrive at a better formulation consider Brownian motion. One may actually observe a large number of Brownian particles and average the result; that means that one really has a physical realization of the ensemble (provided the particles do not interact). However, more often one observes one and the same particle on successive time intervals. The result will be the same if one assume that trajectories that lie time intervals apart are statistically independent. In practice one simply observes the trajectory of a single particle during a long time. The idea is that the irregularly varying function may be cut into a collection of long time intervals and that this collection can serve as the ensemble that defines the stochastic process. The condition for this “self-averaging” to work is that the behaviour of the function during one interval does not affect the behaviour during the next time interval. If this is so the time average equals the ensemble average and the process is called ergodic. This justification by means of self-averaging applies to stationary cases, because one may then choose the time intervals as long as one wishes. If the state of the system changes with time one must choose the interval long enough to smooth out the rapid fluctuations, but short compared to the overall change. The basic assumption for ergodic processes is that such an intermediate interval exists.

2.2 Markov processes and the Chapman-Kolmogorov equation

A Markov process is defined by the following relation, which is called the Markov property

\[
P_{1[n-1]}(x_n, t_n | x_{n-1}, t_{n-1}; \ldots; x_1, t_1) = P_{1|1}(x_n, t_n | x_{n-1}, t_{n-1}), \quad t_1 < t_2 < \ldots < t_n \quad (2.15)
\]

The Markov property expresses that for a Markov process the probability of transition from the value \( x_{n-1} \) that the stochastic variable \( X \) takes at \( t_{n-1} \) to the value \( x_n \) taken at \( t_n \), depends only on the value of \( X \) at the time \( t_{n-1} \) and not on the previous history of the system. In other words, if the present state of the system is known, we can determine the probability of any future state without reference to the past. We say that Markov processes have “short” memory. \( P_{1|1} \) is called the transition probability. The conditional probability \( P_{1|1} \) must satisfy the following properties:

\[
P_{1|1} \geq 0 \quad \text{(non negative)} \quad (2.16)
\]

\[
\int dx_2 P_{1|1}(x_2, t_2 | x_1, t_1) = 1 \quad \text{(at } t_2 \text{ the particle must be somewhere)} \quad (2.17)
\]

\[
P_1(x_2, t_2) = \int dx_1 P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \quad (2.18)
\]

Property (2.18) follows from Bayes relation \( P_2(x_2, t_2; x_1, t_1) = P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \) when integrated over \( x_1 \), and using the fact that \( P_1(x_2, t_2) = \int dx_1 P_2(x_1, t_1; x_2, t_2) \) is the marginal probability distribution of \( P_2 \) with respect to \( x_2 \).
2.3. THE MASTER EQUATION

For a Markov process the joint probabilities for \( n \geq 3 \) are all expressed in terms of \( P_1 \) and \( P_{1|1} \). For \( n = 3 \) we have

\[
P_3(x_1, t_1; x_2, t_2; x_3, t_3) = P_2(x_1, t_1; x_2, t_2) P_{1|2}(x_3, t_3|x_1, t_1; x_2, t_2)
= P_1(x_1, t_1) P_{1|1}(x_2, t_2|x_1, t_1) P_{1|1}(x_3, t_3|x_2, t_2)
\]

(2.19)

For general \( n \)

\[
P_n(x_1, t_1; \ldots; x_n, t_n) = P_{1|1}(x_n, t_n|x_{n-1}, t_{n-1}) \ldots P_{1|1}(x_2, t_2|x_1, t_1) P_1(x_1, t_1)
\]

(2.20)

Taking relation (2.19), integrating it over \( x_2 \) and dividing both sides by \( P_1 \) gives us the Chapman-Kolmogorov (C-K) equation \((t_1 < t_2 < t_3)\)

\[
P_{1|1}(x_3, t_3|x_1, t_1) = \int dx_2 P_{1|1}(x_3, t_3|x_2, t_2) P_{1|1}(x_2, t_2|x_1, t_1)
\]

(2.21)

This equation states that a process starting at \( t_1 \) with value \( x_1 \) reaches \( x_3 \) at \( t_3 \) via any one of the possible values \( x_2 \) at the intermediate time \( t_2 \).

Because Markov processes are fully specified by \( P_1 \) and \( P_{1|1} \), (the whole hierarchy \( P_n \) can be constructed from them), a Markov process is stationary if \( P_1 \) is independent on time and the transition probability \( P_{1|1}(x_2, t_2|x_1, t_1) \) depends only on the time interval \( t_2 - t_1 \), i.e.

\[
P_{1|1}(x_2, t_2|x_1, t_1) = P_{1|1}(x_2, \tau|x_1, 0), \quad \tau = t_2 - t_1
\]

(2.22)

A homogeneous process is a non-stationary Markov process for which the transition probability \( P_{1|1} \) only depends on time differences, i.e. it is given by (2.22).

For stationary and homogeneous processes, a special notation is used for the transition probability

\[
P_{1|1}(x_2, t_2|x_1, t_1) = T_\tau(x_2|x_1), \quad \tau = t_2 - t_1
\]

(2.23)

and the C-K equation reads \((\tau, \tau' > 0)\)

\[
T_{\tau+\tau'}(x_3|x_1) = \int dx_2 T_\tau(x_3|x_2) T_{\tau'}(x_2|x_1),
\]

(2.24)

Exercise Explain why equation (2.21) does not hold for non-Markovian processes.

2.3 The Master equation

The Chapman-Kolmogorov equation for Markov processes is not of much assistance when one searches for solutions of a given problem, because it is essentially a property of the solution. However it can be cast into a more useful form, the Master equation. This is a differential equation for the transition probability. In order to derive it, one first needs to ascertain how the transition probability behaves for short time differences.
Firstly, on inspecting the Chapman-Kolmogorov equation for equal time arguments one find the natural result

\[ P(x_3, t_3 | x_1, t) = \int dx_2 P(x_3, t_3 | x_2, t) P(x_2, t | x_1, t) \rightarrow P(x_2, t | x_1, t) = \delta(x_2 - x_1) \quad (2.25) \]

which is the zero-th order term in the short time behaviour of \( P(x', t' | x, t) \), i.e. for \( t' - t \) small. Keeping this in mind we assume that for small \( \tau \) the transition probability \( P(x_2, t + \tau | x_1, t) \) can be Taylor expanded as follows:

\[ P(x_2, t + \tau | x_1, t) = \delta(x_1 - x_2) + \tau W_t(x_2 | x_1) + O(\tau^2) \quad (2.26) \]

The delta function expresses that the probability to stay at the same state after a time interval equal to zero is one, whereas the probability to change state after a time interval equal to zero is zero. \( W_t(x_2 | x_1) \) is the time derivative of the transition probability at \( \tau = 0 \) and is interpreted as the transition probability per unit time from \( x_1 \) to \( x_2 \) at time \( t \).

This expression must satisfy the normalization property. Therefore, the integral over \( x_2 \) must equal one. In order for this to happen, the above form must be corrected in the following sense:

\[ P(x_2, t + \tau | x_1, t) = \left[ 1 - \tau a^{(0)}(x_1, t) \right] \delta(x_1 - x_2) + \tau W_t(x_2 | x_1) + O(\tau^2) \quad (2.27) \]

where

\[ a^{(0)}(x_1, t) = \int dx_2 W_t(x_2 | x_1) \quad (2.28) \]

The delta function has been corrected by the coefficient \( 1 - \tau \int dx_2 W(x_2 | x_1) \) which corresponds to the probability for no transition to have taken place at all.

From the definition of conditional probability and marginals we have

\[ P_2(x_1, t_1; x_2, t_2) = P_1(x_1, t_1) P_{1|1}(x_2, t_2 | x_1, t_1) \quad (2.29) \]

and

\[ P_1(x_2, t_2) = \int dx_1 P_1(x_1, t_1) P_{1|1}(x_2, t_2 | x_1, t_1) \quad (2.30) \]

In order to find a differential equation for the probability density \( P_1(x, t) \) we rewrite the above equation in the form

\[ P_1(x, t + \tau) = \int dx' P_1(x', t) P_{1|1}(x, t + \tau | x', t) \quad (2.31) \]

The time derivative of \( P_1(x, t) \) is

\[ \frac{\partial P_1(x, t)}{\partial t} = \lim_{\tau \rightarrow 0} \frac{P_1(x, t + \tau) - P_1(x, t)}{\tau} \quad (2.32) \]

and therefore to take the time derivative of (2.31) we need to evaluate the quantity

\[ \lim_{\tau \rightarrow 0} \frac{P_{1|1}(x, t + \tau | x', t) - P_{1|1}(x, t | x', t)}{\tau} = W(x | x') - \int dx W(x | x') \delta(x - x') \quad (2.33) \]

where in the last equality we used (2.27).
2.3. THE MASTER EQUATION

Inserting into (2.32) gives us the differential form of the CK equation which is called *Master equation*

\[
\frac{\partial}{\partial t} P_1(x,t) = \int dx' [W(x|x')P_1(x',t) - W(x'|x)P_1(x,t)]
\] (2.34)

This gives the rate of change of the probability density \(P_1(x,t)\) due to transitions into the state \(x\) from all the other state \(x'\) and transitions out of state \(x\) into all other states \(x'\). There are many applications of the master equation and we shall consider some of them in the subsequent sections. Note that in the Master equation \(t\) enters linearly in the first derivative. The master equation is not invariant under time reversal. This equation describes the irreversible behaviour of the system.

Given a Markov process \(X(t)\), defined by \(P_1(x_1,t_1)\) and \(P_{1\mid1}(x_2,t_2|x_1,t_1)\), we can always extract a sub-ensemble of \(X(t)\), \(X^\star(t)\), characterised by \(P_1^\star(x_1,t_1) = P(x_1,t_1|x_0,t_0)\) and \(P_{1\mid1}^\star(x_2,t_2| x_1,t_1) = P_{1\mid1}(x_2,t_2|x_1,t_1)\), i.e. by taking the sharp value \(x_0\) at \(t_0\), since \(P_1^\star(x_1,t_0) = \delta(x_1 - x_0)\). So the master equation is an equation for all transition probabilities \(P_{1\mid1}(x,t|x_0,t_0)\).

If the range of \(X\) is a discrete set of states with labels \(n\), the equation reduces to

\[
\frac{dp_n(t)}{dt} = \sum_{n'}[W_{nn'}p_{n'}(t) - W_{n'n}p_n(t)]
\] (2.35)

The physical meaning of the master equation is now clear: it is an equation of motion for the probability \(p_n\) of a state \(n\). It is essentially a gain-loss equation: the first term is the gain for the probability of state \(n\) due to transitions from other states \(n'\), and the second term is the loss due to transitions into other states \(n'\).

Owing to \(W(x|x')\tau\) is the transition probability in a short time interval \(\tau\), it can be computed, for the system under study, by means of any available method valid for short times. Then, the master equation serves to determine the evolution of the system over long time periods, at the expense of assuming the Markov property.

The Markov equation can readily be extended to the case of a multicomponent Markov process \(X_i(t), i = 1,2,...,N\) on noting that the Chapman-Kolmogorov equation is valid as it stands by merely replacing \(x\) by \(x = (x_1,...,x_N)\).

2.3.1 Steady state

For a stationary or steady state solution, the left hand side of the Master equation equals zero. Therefore the total number of transition per time into state \(n\) must balance the total number of transition per time out of state \(n\), i.e. we have the balance for *steady state*

\[
\sum_{n'}W_{nn'}p_{n'} = \left(\sum_{n'}W_{n'n}\right)p_n
\] (2.36)

One has *detailed balance* if each individual transition is balanced, i.e. if the number of transitions per time from state \(n\) into state \(n'\) balances the number of transitions per time from \(n'\) to \(n\)

\[
W_{nn'}p_{n'} = W_{n'n}p_n
\] (2.37)
When detailed balance holds, the stationary distribution is often called an equilibrium distribution. Note however that detailed balance is a necessary but not sufficient condition for thermodynamic equilibrium. Detailed balance in thermodynamic equilibrium follows from microscopic time reversibility. However, it may also hold in some cases for non-equilibrium.

### 2.3.2 Convergence to equilibrium

Here we study the conditions under which the dynamics given by the master equation (2.34) converges to a unique stationary (equilibrium) distribution

\[ p_n(\infty) = \frac{1}{Z} e^{-\beta H(n)} \quad (2.38) \]

We show that if the transition rates are in detailed balance with a candidate distribution of the form (2.38), then this is a sufficient, although not necessary, condition for the probability \( p_n(t) \) to converge to \( p_n(\infty) \). In addition, if the process is ergodic, \( p_\infty \) is the unique stationary (equilibrium) distribution and the process always evolve to \( p_n(\infty) = e^{-\beta H(n)}/Z \).

The physically most intuitive convergence proofs are based on constructing a Lyapunov function. Note however that detailed balance is a necessary but not sufficient condition for thermodynamic equilibrium follows from microscopic detailed balance in thermodynamic equilibrium.

We now use the detailed balance condition \( W_{nn'} e^{-H(n')} = W_{n'n} e^{-H(n)} \)

\[
\frac{dF}{dt} = -\frac{1}{2} \sum_{n' n} W_{nn'} e^{-H(n')} [\ln p_n(t) + H(n)] - (\ln p_{n'}(t) + H(n')) [W_{nn'} p_{n'}(t) - W_{n'n} p_n(t)] 
\]

\[
\times [W_{nn'} e^{-H(n')} e^{H(n')} + \ln p_{n'}(t) - W_{n'n} e^{-H(n)} e^{H(n)} + \ln p_n(t)] \leq 0 \quad (2.41)
\]

The last step derives from the general identity \( (e^x - e^y)(x - y) \geq 0 \) for all \( (x, y) \) with equality only if \( x = y \). Since \( F(t) \) is bounded from below, it is indeed a Lyapunov function for the process.
The distance between $p_n(t)$ and the measure $p_n(\infty)$ decreases monotonically until $dF/dt = 0$, i.e. until $p_n(t)$ has reached a point $p_n$ such that

$$e^{H(n)}p_n = e^{-H(n')}p_{n'} \text{ or } W_{nn'} = 0 \quad (2.42)$$

So either $p_n = \chi(n)e^{\beta H(n)}$, with $\chi(n) = \chi(n')$ for all $n,n'$, or $W_{nn'} = 0$ for all $n,n'$. If we define $S_n$ the set of states that are dynamically accessible from $n$, then for all $n' \in S_n$ we need to have $\chi(n) = \chi(n')$. One such stationary solution $p_n$ is the equilibrium measure, corresponding to $\chi(n) = Z^{-1}$ for all $x$. If our process is ergodic, i.e. if $S_n$ coincides with the full range of $n$, then it is the only such distribution to satisfy (2.42) and our process must always evolve towards Boltzmann distribution $p_n(\infty)$.

### 2.3.3 Mean-field equation

From the Master equation one can derive the dynamical equations for the averages of a Markov stochastic process. Let $X$ be a physical quantity with a Markov character. The master equation determinates its probability distribution at all $t > 0$. In ordinary macroscopic physics, however, one ignores fluctuations and treats $X$ as if it were a non-stochastic, single valued quantity $X$. The evolution of $X$ is described by a deterministic differential equation for $X$ called the mean field equation. As the master equation determines the entire probability distribution, it must be possible to derive from it the mean field equation as an approximation for the case that fluctuations are negligible. Let us first write the equation for the time evolution of an arbitrary function $\langle f(X) \rangle$:

$$\frac{d}{dt}\langle f(X) \rangle = \int f(x) \frac{\partial P_1(x,t)}{\partial t} dx = \int f(x)[W(x|x')P_1(x',t) - W(x'|x)P_1(x,t)]dxdx'$$

$$= \int [f(x') - f(x)]W(x'|x)P_1(x,t)dxdx' \quad (2.43)$$

On applying this to $f(X) = X$ one has

$$\frac{d\langle x \rangle}{dt} = \int dx a_1(x)P_1(x,t) = \langle a_1(x,t) \rangle \quad (2.44)$$

with

$$a_\nu(x,t) = \int dx' (x' - x)\nu W_1(x'|x) \quad \nu = 0,1,... \quad (2.45)$$

Note that when $a_1$ is linear function of $x$, one has $\langle a_1(x,t) \rangle = a_1(\langle x \rangle,t)$, whence the mean field equation for the time evolution of $\langle X(t) \rangle$, is the same as

$$\frac{d}{dt}\langle X \rangle = a_1(\langle X \rangle,t) \quad (2.46)$$

which is an ordinary differential equation and can be identified with the mean-field equation for the system.

If however $a_1(\langle X \rangle)$ is not a linear function of $\langle X \rangle$, (2.44) is not a closed equation for $\langle x \rangle$ but higher order moments enter as well. The evolution of $\langle X \rangle$ in the course of time is therefore not determined by $\langle X \rangle$ itself, but is influenced by the fluctuations around this average (variance...
So for non-linear \( a_1(X) \) we need an equation for the variance as well. Similarly, using \( f(X) = X^2 \),

\[
\frac{d}{dt} \langle X^2 \rangle = \int dx\, dx' \, (x'^2 - x^2) W(x'|x) P_1(x,t) \\
= \int dx\, dx' \, [(x' - x)^2 + 2x(x' - x)] W(x'|x) P_1(x,t) \\
= \langle a_2(X) \rangle + 2\langle Xa_1(X) \rangle 
\]

(2.47)

This is identical with

\[
\frac{d\sigma^2}{dt} = \langle a_2(X) \rangle + 2\langle (X - \langle X \rangle) a_1(X) \rangle 
\]

(2.48)

However, if \( a_2 \) is a non-linear function of \( X \), the equation involves even higher order moments.

### 2.4 Examples and exercises

The following two examples of Markov processes are of fundamental importance.

#### 2.4.1 The Wiener-Levy process

The Wiener-Levy process was originally introduced to describe the behaviour of the position of a free Brownian particle in one dimension. On the other hand, it plays a central role in the rigorous foundation of the stochastic differential equations. The Wiener-Levy process is defined in the range \(-\infty < x < \infty \) through \( (t_2 > t_1 > 0) \)

\[
P_{1|1}(x_2,t_2|x_1,t_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \exp \left[ -\frac{(x_2 - x_1)^2}{2(t_2 - t_1)} \right] 
\]

with \( P_1(X_1,0) = \delta(x_1) \) (2.49)

Then, the probability density for \( t_1 > 0 \) is, according to (2.30),

\[
P_1(x_1,t_1) = \frac{1}{\sqrt{2\pi t_1}} \exp \left( -\frac{x_1^2}{2t_1} \right) 
\]

(2.51)

This is a non-stationary \( (P_1 \text{ depends on } t_1) \), Gaussian process. (A process is called a Gaussian process if all its \( P_n \) are multivariate Gaussian distributions. In that case all cumulants \( \kappa_m \) beyond \( m = 2 \) are zero. Gaussian processes are often used to approximate physical processes where it can be assumed that higher order cumulants are negligible).

**Exercise** Show that

\[
\langle X(t_1) \rangle = 0 \quad \text{and} \quad \langle X(t_1)X(t_2) \rangle = \min(t_1,t_2) 
\]

(2.52)

#### 2.4.2 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process was constructed to describe the behaviour of the velocity of a free Brownian particle in one dimension. It also describes the position of an overdamped particle
in a harmonic potential. It is defined by \( \tau = t_2 - t_1 > 0 \)

\[
P_1(x_1) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{x_1^2}{2}\right]
\]

\[
P(x_2, t_2|x_1, t_1) = \frac{1}{\sqrt{2\pi(1-e^{-2\tau})}} \exp \left[-\frac{(x_2 - x_1e^{-\tau})^2}{2(1-e^{-2\tau})}\right]
\]

(2.53)

The O-U process is stationary, Gaussian and Markovian. According to Doob theorem, it is essentially the only process with these three properties.

(a) The Gaussian property is clear for \( P_1 \). By using \( P_2(x_2, t_2; x_1, t_1) = P_1(x_1)P_{1|1}(x_2, t_2|x_1, t_1) \) show that \( P_2(x_2, t_2; x_1, t_1) \) can be identified with the bivariate Gaussian distribution (1.50) with the following parameters

\[
\lambda = e^{-\tau}, \quad \sigma^2 = 1 - e^{-2\tau}
\]

with the particularity that \( \sigma^2 = 1 - \lambda^2 \) in this case.

(b) Show that the O-U process has an exponential autocorrelation function

\[
\langle X(t_1)X(t_2) \rangle = e^{-\tau}
\]

(2.55)

The evolution with time of the distribution \( P_2(x_2, t_2; x_1, t_1) \), seen as the velocity of a Brownian particle, has a clear meaning. For short time differences the velocity is strongly correlated with itself: \( \lambda \approx 1 \) and the variance \( \sigma^2 \) of the distribution shrinks to zero. As time elapses, \( \lambda \) decreases and for long time differences \( \lambda \approx 0 \) and the velocity has lost all memory of its value at the initial time due to the collisions and hence \( P_2(x_2, t_2; x_1, t_1) \) is completely uncorrelated.

**Proof of Doob theorem** Next we show that if \( X(t) \) is a stationary Gaussian Markov process then \( X(t) \) is Ornstein-Uhlenbeck.

Let \( X(t) \) be a stationary Gaussian Markov process. By shifting and rescaling we can always ensure that \( P_1(x) \) is Gaussian with zero mean and unit variance, i.e.

\[
P_1(x_1) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{x_1^2}{2}\right]
\]

(2.56)

The transition probability is Gaussian and has therefore the general form

\[
P(x_2, t_2|x_1, t_1) = De^{-\frac{1}{2}(Ax_2^2+2Bx_1x_2+Cx_1^2)}
\]

(2.57)

where \( A, B, C, D \) are functions of \( \tau \). The normalization yields \( D = \sqrt{A/2\pi} \) and \( C = B^2/A \) so

\[
P(x_2, t_2|x_1, t_1) = \sqrt{\frac{A}{2\pi}} e^{-\frac{1}{2}A(x_2 + \frac{Bx_1}{A})^2}
\]

(2.58)

Using \( \int dx_1 P_{1|1}(x_2, t_2|x_1, t_1)P_1(x_1, t_1) = P(x_2, t_2) \) gives in addition \( B^2 = A(A-1) \). The one remaining unknown parameter \( A \) can be expressed in the equally unknown correlation function using

\[
\kappa(t_2 - t_1) = \int dx_1 dx_2 x_1x_2 P_{1|1}(x_2, t_2|x_1, t_1)
\]

(2.59)
which yields $A = (1 - \kappa^2(t_2 - t_1))^{-1}$. Hence,

$$P(x_2, t_2| x_1, t_1) = \frac{1}{\sqrt{2\pi(1 - \kappa^2(t_2 - t_1))}} e^{-\frac{(x_2 - \kappa(t_2 - t_1)x_1)^2}{2(1 - \kappa^2)}}$$

(2.60)

Now take a third time $t_3$ and use (2.21) in (2.59)

$$\kappa(t_3 - t_1) = \int dx_1 x_1 P_1(x_1| x_2, t_2| x_1, t_1) P_1(x_1)$$

$$= \int dx_1 dx_2 \kappa(t_3 - t_2)x_2x_1 P_1(x_2, t_2| x_1, t_1) P_1(x_1) = \kappa(t_3 - t_2)\kappa(t_2 - t_1)$$

(2.61)

The functional relation for $\kappa(\tau)$ shows that

$$\kappa(\tau) = e^{-\gamma\tau}$$

(2.62)

Substitution of this function in (2.60) gives (2.53) and completes the proof.

**Exercise** Prove, using the generating functional or otherwise, that if $X(t)$ is stationary, Gaussian (with zero mean and unit variance) and has an exponential correlation function $\kappa(\tau) = \kappa(0)e^{-\gamma\tau}$, then $X(t)$ is Ornstein-Uhlenbeck and hence Markovian.

### 2.4.3 The random walk in one dimension

A man moves along a line, taking, at random, steps to the left or to the right with equal probability. The steps are of length $d$ so that his position can take on only the values $nd$, where $n$ is integral.

The problem can be defined in two ways, leading to discrete time and continuous time random walks. The discrete time, which is more traditional, is to allow the walker to take steps at times $n\tau$ ($n$ integral) at which time he must step either left or right, with equal probability, so that

$$P_{1|1}(nd, s\tau| md, (s-1)\tau) = \frac{1}{2}\delta_{n,m-1} + \frac{1}{2}\delta_{n,m+1}$$

(2.63)

**Exercise** Show that in the limit of small jumps, and continuous time, the random walk is described by a Wiener process.

**Answer** We want to find an equation for the evolution of the probability density to find the walker at a certain distance after a given time is elapsed and show that this is solved by the Wiener process.

From (2.30) we have

$$P_1(nd, s\tau) = \sum_{n'} P_1(nd, s\tau| n'd, (s-1)\tau)P_1(n'd, (s-1)\tau) = \frac{1}{2}P_1((n-1)d, (s-1)\tau) + \frac{1}{2}P_1((n+1)d, (s-1)\tau)$$

(2.64)
and therefore

\[ P_1(nd, s \tau) - P_1(nd, (s - 1) \tau) = \frac{1}{2} [ P_1((n-1)d, (s-1)\tau) + P_1((n+1)d, (s-1)\tau) - 2P_1(nd, (s-1)\tau)] \]

(2.65)

We rewrite equation (2.65)

\[ \frac{P_1(nd, s \tau) - P_1(nd, (s - 1) \tau)}{\tau} = \frac{P_1((n-1)d, (s-1)\tau) + P_1((n+1)d, (s-1)\tau) - 2P_1(nd, (s-1)\tau)}{d^2} \frac{d^2}{\tau} \]

(2.66)

If we now set \( nd = x \), \( (s-1)\tau = t \) and take the limit \( d \to 0 \), \( \tau \to 0 \) in such a way that \( D = \frac{d^2}{2\tau} \) is finite, we obtain

\[ \partial P_1(x, t) \partial t = D \lim_{d \to 0} \left[ \frac{P_1(x-d, t) - P_1(x, t)}{d^2} + \frac{P_1(x+d, t) - P_1(x, t)}{d^2} \right] \]

\[ = D \lim_{d \to 0} \left[ \frac{1}{d} \frac{\partial P_1(x, t)}{\partial x} - \frac{1}{d} \frac{\partial P_1(x-d, t)}{\partial x} \right] \]

\[ = D \lim_{d \to 0} \frac{\partial}{\partial x} \left( P_1(x, t) - P_1(x-d, t) \right) = D \frac{\partial^2 P_1(x, t)}{\partial x^2} \]

(2.67)

So we have obtained the Fokker-Planck equation for diffusion

\[ \frac{\partial P_1(x, t)}{\partial t} = D \frac{\partial^2 P_1(x, t)}{\partial x^2} \]

(2.68)

Let us now solve equation (2.68) assuming that, initially, \( P_1(x, 0) = \delta(x) \), where \( \delta(x) \) is a Dirac delta function. We first introduce the Fourier transform of \( P(x, t) \):

\[ P(k, t) = \int_{-\infty}^{\infty} dx \ P_1(x, t) e^{ikx} \]

(2.69)

Then the diffusion equation takes the form

\[ \frac{\partial P(k, t)}{\partial t} = -Dk^2 P(k, t) \]

(2.70)

We can solve this to obtain

\[ P(k, t) = P_1(k, 0) e^{-Dk^2 t} \]

(2.71)

The solution contains an arbitrary function \( P_1(k, 0) \), which can be determined from the initial condition, as \( P_1(k, 0) = 1 \). We can now take the inverse transform to obtain

\[ P_1(x, t|0, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \ e^{-ikx} e^{-Dk^2 t} = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \]

(2.72)

This gives the probability of finding the particle at point \( x \) at time \( t \) if it started at \( x = 0 \) at time \( t = 0 \) and corresponds to the Wiener process. It is interesting to obtain the equation of motion of the moments \( \langle x(t) \rangle \) and \( \langle x^2(t) \rangle \). The moment

\[ \langle x(t) \rangle = \int_{-\infty}^{\infty} dx \ x P_1(x, t) \]

(2.73)
gives the average position of the particle as a function of time. If we multiply (2.68) by \(x\) and integrate over \(x\), by carrying an integration by parts and using the boundary conditions \(P_1(\pm \infty, t) = 0\), we find
\[
\frac{d\langle x(t) \rangle}{dt} = 0
\] (2.74)
Thus, the average position of the particle does not change with time.

Let us now consider the moment \(\langle x^2(t) \rangle\). If we multiply (2.68) by \(x^2\) and integrate over \(x\) we find, by performing two integrations by part and using the boundary conditions \(P_1(\pm \infty, t) = 0\) and \(\partial P_1(\pm \infty, t)/\partial x = 0\),
\[
\frac{d\langle x^2(t) \rangle}{dt} = 2D
\] (2.75)
or
\[
\langle x^2(t) \rangle = 2Dt
\] (2.76)
This is characteristic of a diffusion process. The moments \(\langle x^2(t) \rangle\) is a measure of the width of \(P_1(x, t)\) at time \(t\). We see that \(P_1(x, t)\) spreads with time, as we expect.
Chapter 3

Markov Chains and One step processes

One of the simplest application of the Master equation is to the case of Markov chains. These are processes which involve transitions between discrete stochastic variables at discrete times.

It is sometimes possible to solve the master equation if we can find the eigenvectors of the transition matrix. Since this is not symmetric in general, the solution involves the introduction of left and right eigenvectors. We shall obtain the general form of solution for the case of homogeneous Markov processes and write it in terms of left and right eigenvectors of the transition matrix.

Another simple application of the Master equation is to one-step processes. These are continuous time Markov processes in discrete state space whose transitions are only between adjacent states. For some one-step processes the master equation can be solved exactly by using a generating function (similar to characteristic function). We can obtain a partial differential equation for the generating function which may be solved in simple cases.

3.1 Markov Chains

One of the simplest example of a Markov process is that of a Markov chain. This involves transitions between values of a discrete stochastic variable $X$ occurring at discrete times. We will limit for simplicity to Markov chains with finite state space $S$. This is not essential, but removes distracting technical complications. Let us assume that $X$ can take on values in $S = (1, 2, ..., N)$ and time is measured in discrete steps $t = s\tau$, where $\tau$ is a fundamental time interval. Then for one step from $t_1 = m\tau$ to $t_2 = n\tau$, $(n > m)$, equation (2.18) can be written as

$$P_1(i, n) = \sum_{j=1}^{N} P_1(j, m) P_{1|1}(i, n|j, m)$$

(3.1)

The theory of Markov chains is most highly developed for homogeneous chains and we shall mostly be concerned with these. A Markov chain is said to be homogeneous when the probabilities $P_{1|1}(i, n|j, m)$ depends on the time interval $n - m$ but not on the time $m$. For such a chain
we define the \( n \)-step transition probabilities

\[
Q_{ji}^{(n)} = P_{1|1}(i, m + n|j, m) = P_{1|1}(i, n|j, 0) \tag{3.2}
\]

Of particular importance are the one-step transition probabilities, which we write simply as \( Q_{ji} \), (they depend only on the states and not on the times),

\[
Q_{ji} = Q_{ji}^{(1)} = P(i, m + 1|j, m) \tag{3.3}
\]

Since the system has to move to some state from any state \( j \) we have, for all \( j \)

\[
\sum_{k=1}^{N} Q_{jk} = 1 \tag{3.4}
\]

The matrix \( Q \) of transition probabilities

\[
Q = \begin{pmatrix}
Q_{11} & Q_{12} & \cdots \\
Q_{21} & Q_{22} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix} \tag{3.5}
\]

is called stochastic matrix, the defining property of a stochastic matrix being that its elements are non-negative and that all its row sums are unity. Note we use the convention that \( Q_{ji} \) denotes the probability of a transition from state \( j \) (row suffix) to \( i \) (column suffix).

Often one is interested in the probability that after \( n \) steps the Markov chain is in a given state. Next we show that this problem reduces to calculating entries in the \( n \)-th power of the transition matrix \( Q \). Let us introduce the vector of state occupation probabilities at time \( n \),

\[
p^{(n)} = (p_1^{(n)}, p_2^{(n)}, ..., p_N^{(n)}) \tag{3.6}
\]

whose components are defined as

\[
p_j^{(n)} = P_j(j, n) \tag{3.7}
\]

We can write equation (2.18), for one step from \( t_1 = n - 1 \) to \( t_2 = n \) as

\[
p_i^{(n)} = \sum_{j=1}^{N} p_j^{(n-1)} Q_{ji} \tag{3.8}
\]

In matrix notation

\[
p^{(n)} = p^{(n-1)} Q \tag{3.9}
\]

and on iteration we obtain

\[
p^{(n)} = p^{(0)} Q^n \tag{3.10}
\]

with \( Q^0 \) being the identity matrix, i.e. \( Q_{ij}^0 = \delta_{ij} \). Consequently, the process can be described completely by giving its initial probability vector \( p^{(0)} \) and the transition matrix \( Q \). The probability of the system taking a specific path of states is

\[
P_T(i_0, 0; ..., i_T, T) = p_{i_0}^{(0)} \prod_{n=1}^{T} Q_{i_n-1, i_n} \tag{3.11}
\]
3.1. MARKOV CHAINS

If the system starts in any given state \( j \), then
\[
p_k^{(0)} = \delta_{kj}
\]
and
\[
p_k^{(n)} = (Q^n)_{jk}
\]
From the fact that \( Q^{m+n} = Q^m Q^n \) we have
\[
Q^{(m+n)}_{jk} = \sum_{\ell=1}^N Q^{(m)}_{j\ell} Q^{(n)}_{\ell k}
\]
which is the Chapman-Kolmogorov equation for homogeneous Markov chain.

In the non-homogeneous case the transition probability
\[
P_{ij|1}(i,s|j,r) \quad (s > r)
\]
will depend on both \( r \) and \( s \). In this case we write
\[
Q_{ij}(r,s) = P_{ij|1}(i,s|j,r)
\]
In particular the one-step transition probabilities \( Q_{ij}(r,r+1) \) will depend on the time \( r \) and we will have a sequence of stochastic matrices
\[
Q(r) = \{Q_{ij}(r,r+1)\} \quad (r = 0, 1, ...)
\]
and the Chapman-Kolmogorov equation reads
\[
Q_{jk}(r,t) = \sum_{\ell} Q_{j\ell}(r,s) Q_{\ell k}(s,t) \quad (r < s < t)
\]
We define stationary distribution of the Markov chain (3.9) a vector \( \pi = (\pi_1, \ldots, \pi_N) \) such that
\[
\pi = \pi Q \quad \text{and} \quad \forall i \in S : \pi_i \geq 0, \quad \sum_j \pi_j = 1
\]
Thus \( \pi \) is a left eigenvector of the stochastic matrix, with eigenvalue \( \lambda = 1 \) and with non-negative entries, and represents a time-independent solution of the Markov chain.

For later reference, we note that the fact that all the row sums of a stochastic matrix are unity can be expressed in matrix notation by
\[
QI = I
\]
with \( I = (1, \ldots, 1) \). So \( I \) is always a right eigenvector of \( Q \), with eigenvalue \( \lambda = 1 \). Pre-multiplying each side by \( Q \) we have \( Q^2 I = QI = I \) and in general
\[
Q^n I = I
\]
so \( Q^n \) is also a stochastic matrix, for each integral \( n \). Finally, we mention that stochastic matrices map normalized probabilities onto normalized probability, as they conserve sign and normalization of the state probabilities. (Prove this statement as an exercise.)
3.1.1 Eigenvalues and eigenvectors of stochastic matrices

We have seen that calculating the probability that the Markov chain is in a given state after \(n\) steps, normally involves calculating entries in the \(n\)-th power of the transition matrix. This is best done by using a spectral representation of the transition matrix, i.e. a decomposition of the matrix based on eigenvalues and eigenvectors. For \(p^n\) to remain well-defined for \(n \to \infty\), it is vital that the eigenvalues are sufficiently small. We next present some facts about eigenvectors and eigenvalues of stochastic matrices.

The matrix \(Q\) is in general not a symmetric matrix, therefore the right and left eigenvectors will be different. We write the left and the right eigenvector problems as

\[
\begin{align*}
\mathbf{x}(j) & = \lambda_j \mathbf{x}(j) \\
\mathbf{y}^{(i)} Q & = \lambda_i \mathbf{y}^{(i)}
\end{align*}
\]

\(\mathbf{x}(j), \mathbf{y}^{(i)} \in \mathbb{C}^N\) and \(\lambda_{i,j} \in \mathbb{C}\) (since \(Q\) need not be symmetric, eigenvalues and eigenvectors need not be real-valued). Much can be extracted from the two defining properties \(Q_{ik} \geq 0 \forall (i,k) \in S^2\) and \(\sum_k Q_{ik} = 1 \forall k \in S\) alone. Note that eigenvectors \((\mathbf{x}, \mathbf{y})\) of \(Q\) need not be probabilities in the sense of the \(p^n\), as they could have negative or complex entries.

- The spectra of left- and right- eigenvalues of \(Q\) are identical.
  
  proof:
  Equations (3.21) and (3.22) give
  \[
  \begin{align*}
  \det[Q - \lambda_R I] & = 0 \\
  \det[Q^\dagger - \lambda_L I] & = 0
  \end{align*}
  \]
  Rewriting gives
  \[
  0 = \det[Q^\dagger - \lambda_L I] = \det[(Q - \lambda_L I)^\dagger] = \det[Q - \lambda_L I]
  \]
  Equating (3.23) and (3.25), which are both equal zero for arbitrary \(Q\), gives \(\lambda_R = \lambda_L = \lambda\).

- Right and left eigenvectors are biorthogonal.
  
  proof:
  Pre-multiplying (3.21) by \(y^{(i)}\), post-multiplying (3.22) by \(x^{(j)}\) and subtracting, we find
  \[
  (\lambda_i - \lambda_j)y^{(i)}x^{(j)} = 0
  \]
  so \(y^{(i)}x^{(j)} = 0\) if \(\lambda_i \neq \lambda_j\), while \(y^{(i)}x^{(i)} \neq 0\). We can scale these vectors so that \(\sum_m y_m^{(i)} x_m^{(j)} = \delta_{ij}\) and (equivalently) \(\sum_i y_i^{(i)} x_i^{(i)} = \delta_{mn}\).

- One can expand the matrix \(Q\) in terms of its left and right eigenvector. From (3.22), \(\sum_n y_n^{(i)} Q_{nm} = \lambda_i y_m^{(i)}\), if we multiply by \(x^{(i)}_\ell\) and sum over \(i\), we obtain
  \[
  Q_{\ell m} = \sum_i \lambda_i x^{(i)}_\ell y^{(i)}_m
  \]
  From \(Q^n x^{(j)} = \lambda^n_j x^{(j)}\) and \(y^{(i)} Q^n = \lambda^n_i y^{(i)}\) one also has
  \[
  (Q^n)_{\ell m} = \sum_i \lambda^n_i x^{(i)}_\ell y^{(i)}_m
  \]
3.1. MARKOV CHAINS

• All eigenvalues $\lambda$ of stochastic matrices $Q$ obey
  \[ yQ = \lambda y, \ y \neq 0 : \quad |\lambda| \leq 1 \]  
  (3.29)

proof:
  Consider the left eigenvalue equation, $\lambda y_i = \sum_j Q_{ji}y_j$, take absolute values of both sides, sum over $i \in S$, and use the triangular inequality:
  \[ \sum_i |\lambda y_i| = \sum_i |\sum_j Q_{ji}y_j| \Rightarrow |\lambda| \sum_i |y_i| = \sum_i |\sum_j Q_{ji}y_j| \leq \sum_i \sum_j |Q_{ji}y_j| = \sum_i \sum_j Q_{ji}|y_j| = \sum_j |y_j| \]
  Since $y \neq 0$ we know that $\sum_i |y_i| \neq 0$ and hence $|\lambda| \leq 1$.

• Since $Q$ has a right eigenvector with eigenvalue $\lambda = 1$, namely $(1, ..., 1)$ due to $\sum_j Q_{ij} = 1, \forall i \in S$, $\lambda = 1$ is always an eigenvalue. Therefore, there must exist at least one left eigenvector with $\lambda = 1$.

3.1.2 Definitions based on accessibility of states

• Regular (or ergodic) Markov Chain
  \[ \exists n \geq 0 : (Q^n)_{ij} > 0 \forall i, j \]  
  (3.30)
  Note that if the above holds for some $n \geq 0$, it will hold for any $n' \geq n$. Thus, irrespective of the initial state $i$, after a finite number of iterations there is a non-zero probability for the system to be in any state.

• Existence of paths
  \[ i \rightarrow j : \quad \exists n \geq 0 \text{ such that } (Q^n)_{ij} > 0 \]  
  (3.31)
  \[ i \not\rightarrow j : \quad \nexists n \geq 0 \text{ such that } (Q^n)_{ij} > 0 \]  
  (3.32)

• Communicating states
  \[ i \leftrightarrow j : \quad \exists n, m \geq 0 \text{ such that } (Q^n)_{ij} > 0 \text{ and } (Q^m)_{ji} > 0 \]  
  (3.33)
  Given sufficient time we can always get from $i$ to $j$ and from $j$ to $i$.

• Closed set
  any set $C \subseteq S$ of states such that $\forall i \in C, \forall j \notin C \ i \not\rightarrow j$  
  (3.34)
  So no state inside $C$ can ever reach any state outside $C$. In other words, $(Q^n)_{ij} = 0$ for all $n \geq 0$ if $i \in C, j \notin C$.

• Absorbing state
  A closed set with just one element. So if $i$ is an absorbing state
  \[ Q_{ji} = 0 \forall j \neq i; \quad Q_{ii} = 1 \]  
  (3.35)
  one cannot leave this state ever. In other words, a Markov chain has an absorbing state if one or more rows of the transition matrix contain all zeros except for the diagonal element, which must be one.
• **Irreducible set of states**
  This is any set $C \subseteq S$ of states such that
  $$\forall \ i, j \in C \ i \leftrightarrow j$$
  (3.36)
  All states in an irreducible set are connected to each other, in that one can go from any state in $C$ to any other state in $C$ in a finite number of steps.

• **Irreducible Markov chain**
  A Markov chain with the property that the complete set of states $S$ is itself irreducible,
  $$\forall \ i, j \in S \ i \leftrightarrow j$$
  Equivalently, one can go from any state in $S$ to any other state in $S$ in a finite number of steps (a transition matrix with absorbing states cannot be irreducible). Note that all regular Markov chains are irreducible while not all irreducible Markov chains are regular. An example of irreducible chain which is not regular is provided by
  \[
  Q = \begin{pmatrix}
  0 & 1 \\
  1 & 0
  \end{pmatrix}
  \] (3.37)
  We have $Q^{2n+1} = Q$ and $Q^{2m} = I$. However there is not $n$ for which $Q^n$ has all entries non-zero.

### 3.1.3 Convergence to a stationary state

One often wishes to know the state of the system after a large number of transitions $n \to \infty$. Does the system retain some memory of its initial state $p^{(0)}$, or after many transitions does the system proceed to some unique final state independent of the initial state? We shall see that the behaviour of the probability vector $Q^{(n)}$, for large $n$, depends on the structure of the transition matrix $Q$.

A fundamental theorem, *Perron-Frobenius theorem* states that a real square matrix with positive entries has a unique largest eigenvalue and the corresponding vector has strictly positive components. As a consequence, if $Q$ is the stochastic matrix of a regular Markov chain, then it has only one eigenvalue with value $\lambda = 1$. Let us denote it $\lambda_1 = 1$. The right eigenvector associated with the unit eigenvalue is $x^{(1)} = (1, ..., 1)$, and we denote the left eigenvector associated to it by $y^{(1)} = \pi$ (we will show shortly that $\pi$ has all its components non-negative and it is normalized to one so it needs to be the stationary distribution, hence the notation). Then (3.28) can be rewritten as

\[
(Q^n)_{\ell m} = x_\ell^{(1)} y_m^{(1)} + \sum_{j \neq 1} \lambda_j^n x_\ell^{(j)} y_m^{(j)} = \pi_m + \sum_{j \neq 1} \lambda_j^n x_\ell^{(j)} y_m^{(j)}
\] (3.38)

Since $\lambda_j < 1 \ \forall \ j \neq 1$, we can take the limit $n \to \infty$ to obtain

$$\lim_{n \to \infty} (Q^n)_{\ell m} = \pi_m$$ (3.39)

This has many useful consequences:
• The sequence of powers of $Q$ ($Q, Q^2, Q^3, \ldots$) tends to a unique transition matrix, $M$, whose rows $\pi = (\pi_1, \ldots, \pi_N)$ are real and obey $\pi_i \geq 0$ and $\sum_i \pi_i = 1$, given that $M$ is a stochastic matrix (due to $Q^n$ being a stochastic matrix for any $n \geq 0$).

$$\lim_{n \to \infty} Q^n = M = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix} \quad (3.40)$$

• It is easy to show that $QM = M$ and $\pi = p^{(0)}M$.

proof: Consider for example the case $n = 2$. Then

$$QM = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} \pi_1 \sum_j Q_{1j} \\ \pi_1 \sum_j Q_{2j} \end{pmatrix} = \begin{pmatrix} \pi_1 \sum_j Q_{1j} \\ \pi_1 \sum_j Q_{2j} \end{pmatrix} = M \quad (3.41)$$

and

$$p^{(0)}M = (p_1^{(0)}, p_2^{(0)}) \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = (\pi_1, \pi_2) = \pi \quad (3.42)$$

because rows of stochastic matrices add to one.

• The solution of the Markov chain $p^{(n)}$ will also converge to the (unique) fixed probability vector $\pi$

$$\lim_{n \to \infty} p^{(n)}_i = \lim_{n \to \infty} \sum_j p_j^{(0)} Q^{(n)}_{ji} = \sum_j p_j^{(0)} Q^{(n)+1}_{ji} = \pi_i \quad (3.43)$$

• Such limiting vector $\pi = \lim_{n \to \infty} p^{(n)}$, is the stationary state of the Markov chain, i.e.

$$\pi = \pi Q \quad (3.44)$$

proof: This follows from

$$\pi = \lim_{n \to \infty} p^{(0)} Q^n = \lim_{n \to \infty} p^{(0)} Q^{n+1} = \pi Q \quad (3.45)$$

We say that the distribution $\pi$ is invariant for $Q$ or that $\pi$ is the steady-state vector of the Markov chain (in such a solution the probability to find any state will not change with time).

Note that the stationary solution to which the Markov chain converges is unique (independent of the choice made for the $p^{(0)}_i$) because all rows of the matrix $M$ are identical.

### 3.1.4 Time-reversal and detailed balance

For Markov chains, the past and future are independent given the present. This property is symmetrical in time and suggests looking at Markov chains with time running backwards. On the other hand, convergence to stationarity shows behaviour which is asymmetrical in time (a highly organised state such as a point mass decays to a disorganised one, the invariant
distribution; this is an example of entropy increasing). It suggests that if we want complete
time symmetry we must begin at stationarity. A Markov chain at stationarity, run backwards,
is again a Markov chain. The transition matrix may however be different. If, in addition,
the Markov chain satisfies detailed balance, then the direct and reverse chains have the same
transition matrix and the chain is called reversible.

- defn: probability current in Markov chains

The net probability current \( J_{i\rightarrow j} \) from any state \( i \in S \) to any state \( j \in S \), in the stationary
state \( \pi = (\pi_1, \ldots, \pi_N) \) of a Markov chain characterized by the stochastic matrix \( Q = \{Q_{ij}\} \)
and state space \( S \), is defined as

\[
J_{i\rightarrow j} = \pi_i Q_{ij} - \pi_j Q_{ji}
\]

(3.46)

consequences, conventions

1. The current is by definition anti-symmetric under permutation of \( i \) and \( j \):

\[
J_{j\rightarrow i} = \pi_j Q_{ji} - \pi_i Q_{ij} = -\left[\pi_i Q_{ij} - \pi_j Q_{ji}\right] = -J_{i\rightarrow j}
\]

2. Imagine that the chain represents a random walk of a particle, in a stationary state. Since
\( \pi_i \) is the probability to find the particle in state \( i \), and \( Q_{ij} \) the likelihood that it
subsequently moves from \( i \) to \( j \), \( \pi_i Q_{ij} \) is the probability that we observe the particle
moving from \( i \) to \( j \). With multiple particles it would be proportional to the number
of observed moves from \( i \) to \( j \). Thus \( J_{i\rightarrow j} \) represents the net balance of observed
transitions between \( i \) and \( j \) in the stationary state; hence the term 'current'. If
\( J_{i\rightarrow j} > 0 \) there are more transitions \( i \rightarrow j \) than \( j \rightarrow i \); if \( J_{i\rightarrow j} < 0 \) there are more
transitions \( j \rightarrow i \) than \( i \rightarrow j \).

3. Conservation of probability implies that the sum over all currents is always zero:

\[
\sum_{ij} J_{i\rightarrow j} = \sum_{ij} \left[\pi_i Q_{ij} - \pi_j Q_{ji}\right] = \sum_i \pi_i \left(\sum_j Q_{ij}\right) - \sum_j \pi_j \left(\sum_i Q_{ji}\right)
\]

\[
= \sum_i \pi_i - \sum_j \pi_j = 1 - 1 = 0
\]

- defn: detailed balance

A stochastic matrix \( Q \) and a measure \( \pi \) are said to be in detailed balance if

\[
\pi_i Q_{ij} = \pi_j Q_{ji} \quad \forall i, j.
\]

(3.47)

If \( Q \) and \( \pi \) are in detailed balance, then \( \pi \) is invariant for \( Q \).

\[
(\pi Q)_j = \sum_i \pi_i Q_{ij} = \sum_i \pi_j Q_{ji} = \pi_j
\]

(3.48)

consequences, conventions
3.2. EXAMPLES AND EXERCISES ON MARKOV CHAINS

1. Let $X_n$ be a Markov chain with transition matrix $Q$ and stationary distribution $\pi$. We say that $X_n$ is reversible if $X_{N-n}$ is also a Markov chain with invariant distribution $\pi$ and transition matrix $Q$. This is equivalent to say that $Q$ and $\pi$ are in detailed balance.

2. We see that detailed balance represents the special case where all individual currents in the stationary state are zero: $J_{i\rightarrow j} = \pi_i Q_{ij} - \pi_j Q_{ji} = 0$ for all $i, j \in S$.

3. Detailed balance is a stronger condition than stationarity. All regular Markov chains with detailed balance have by definition a unique stationary state, but not all regular Markov chains with stationary states obey detailed balance.

4. Markov chains used to model closed physical many-particle systems with noise are usually of the detailed balance type, as a result of the invariance of Newton’s laws of motion under time reversal $t \rightarrow -t$.

Example Consider the transition matrix

$$Q = \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix} \quad (3.49)$$

The invariant distribution $\pi$ is found from $\pi = \pi Q$ as $\pi = (1/3, 1/3, 1/3)$. Clearly, $\pi_i Q_{ij} \neq \pi_j Q_{ji}$ because $Q$ is not symmetric. Therefore the chain is not reversible.

3.2 Examples and exercises on Markov chains

3.2.1 Two-state Markov chain

Let us consider in some detail a Markov chain with two states. This is the simplest non-trivial state space. The most general two-state chain has transition matrix of the form

$$Q = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad (3.50)$$

In order to calculate the time-dependent probabilities $p^{(n)}$ for given initial probabilities $p^{(0)}$, it is often useful to use the diagonal or spectral representation of $Q$. Suppose that $Q$ has distinct eigenvalues, $\lambda_1, \lambda_2$. Then it is a standard result of matrix theory that we can find a $2 \times 2$ matrix $U$ such that

$$Q = U \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} U^{-1} \quad (3.51)$$

where the columns $q_1, q_2$ of $U$ are solutions of the equations

$$Q q_i = \lambda_i q_i \quad (i = 1, 2) \quad (3.52)$$

Hence we have

$$Q^n = U \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} U^{-1} \quad (3.53)$$
Let us therefore carry out this representation for the two-state Markov chain. The eigenvalues of \( Q \) are the solution of the characteristic equation

\[
|Q - \lambda I| = 0
\]  
(3.54)

Hence, \( \lambda_1 = 1 \) and \( \lambda_2 = 1 - \alpha - \beta \) (and \( \lambda_1 \neq \lambda_2 \), provided \( \alpha + \beta \neq 0 \)). Note also that eigenvalue \( \lambda_2 \) is in modulus less than unity unless \( \alpha + \beta = 0 \) or \( \alpha + \beta = 2 \). The corresponding eigenvectors are

\[
\lambda_1 = 1 : \quad \begin{cases} (1 - \alpha)x_1^1 + \alpha x_2^1 = x_1^1 \\ \beta x_1^1 + (1 - \beta)x_2^1 = x_2^1 \end{cases} \quad \Rightarrow (x_1^1, x_2^1) = (1, 1)
\]  
(3.55)

\[
\lambda_2 = 1 - \alpha - \beta : \quad \begin{cases} (1 - \alpha)x_1^2 + \alpha x_2^2 = (1 - \alpha - \beta)x_1^2 \\ \beta x_1^2 + (1 - \beta)x_2^2 = (1 - \alpha - \beta)x_2^2 \end{cases} \quad \Rightarrow (x_1^2, x_2^2) = (\alpha, -\beta)
\]  
(3.56)

We may take then

\[
U = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \quad \text{so} \quad U^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}
\]  
(3.57)

Hence,

\[
Q = U \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix} U^{-1}
\]  
(3.58)

Note that the rows of \( U^{-1} \) are given by the left eigenvectors \((y_1^1, y_2^1) = (\beta, \alpha)\) and \((y_1^2, y_2^2) = (1, -1)\), associated to \( \lambda_1 = 1 \) and \( \lambda_2 = 1 - \alpha - \beta \), respectively. Also, left and right eigenvectors are biorthogonal, as they should. We can express the solution in terms of the eigenvectors of \( Q \).

One way to do that is to normalize the eigenvectors, so that \( x^{(i)}y^{(j)} = \delta_{ij} \), i.e. introduce \( e^{(1)} = x^{(1)}/(x^{(1)}y^{(1)}) \) and \( e^{(2)} = x^{(2)}/(x^{(2)}y^{(2)}) \), and expand \((Q^n)_{ij} = \sum_{\ell=1}^2 \lambda_{\ell}^n e^{(\ell)}_i y^{(\ell)}_j \). Alternatively, we can perform a matrix multiplication

\[
Q^n = \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - \alpha - \beta)^n \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}
\]  
(3.59)

As we said, one question that often arises is whether after a sufficiently long period of time the system settles down to a condition of statistical equilibrium in which the state occupation probabilities are independent of the initial condition. This is so if the matrix is regular, i.e. for \( \alpha \neq 0 \) and \( \beta \neq 0 \) and \( \alpha + \beta \neq 2 \). The equilibrium distribution is found from \( \pi = \pi Q \)

\[
\alpha \pi_1 - \beta \pi_2 = 0, \quad -\alpha \pi_1 + \beta \pi_1 = 0
\]  
(3.60)

Clearly, \(|I - Q| = 0\), so one of the equation is redundant and another equation is required to fix \( \pi \) uniquely. The extra equation is provided by the fact that we need the condition \( \pi_1 + \pi_2 = 1 \) for a probability distribution, so

\[
\pi_1 = \frac{\beta}{\alpha + \beta} \quad \pi_2 = \frac{\alpha}{\alpha + \beta}
\]  
(3.61)
3.2. EXAMPLES AND EXERCISES ON MARKOV CHAINS

The first term in (3.59) is constant and is seen to be

\[
\left( \frac{\pi_1}{\pi_1} \frac{\pi_2}{\pi_2} \right)
\]

while the second term is a transient term and tends to zero rapidly as \( n \) increases, since \( |1 - \alpha - \beta| < 1 \). Thus, as \( n \to \infty \),

\[
Q^n \to \left( \frac{\pi_1}{\pi_1} \frac{\pi_2}{\pi_2} \right)
\]

and

\[
p^{(n)} \to p^{(0)} \left( \frac{\pi_1}{\pi_1} \frac{\pi_2}{\pi_2} \right) = (\pi_1, \pi_2) = \pi
\]

In the case of two state Markov chains we may use an alternative route to calculate the \( n \)-step transition probability from a given state \( i \) to another state \( j \). We may exploit the relation

\[
Q^{n+1} = Q^n Q
\]

to get a recurrence equation for \( Q^{(n)}_{ij} \). As an example, we write

\[
Q^{(n+1)}_{11} = \beta + (1 - \alpha - \beta) Q^{(n)}_{11}, \quad Q^{(0)}_{11} = 1
\]

This equation has the form of the recurrence relation

\[
x_{n+1} = ax_n + b
\]

The way one solves such a relation is by looking first for a constant solution \( x_n = x \). Then \( x = ax + b \), so provided \( a \neq 1 \) we have \( x = b/(1-a) \). Now \( y_n = x_n - b/(1-a) \) satisfies \( y_n = ay_n \) so \( y_n = a^n y_0 \). Thus the general solution for \( a \neq 1 \) is given by

\[
x_n = Aa^n + \frac{b}{1-a}
\]

where \( A = x_0 - b/(1-a) \) is constant. When \( a = 1 \) the general solution is obviously

\[
x_n = x_0 + nb
\]

Thus equation (3.66) has a unique solution

\[
Q^{(n)}_{11} = \begin{cases} 
\frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1 - \alpha - \beta)^n & \text{for } \alpha + \beta > 0 \\
\frac{1}{\alpha+\beta} & \text{for } \alpha + \beta = 0 
\end{cases}
\]

Exercise (Virus Mutation) - Suppose a virus can exist in \( N \) different strains and in each generation either stays the same or with a probability \( \alpha \) mutates to another strain, which is
chosen at random. What is the probability that the strain in the $n$-th generation is the same as that in the 0-th?

Answer: $\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(1 - \frac{\alpha N}{N-1}\right)^n \quad (3.71)$

**Exercise** (Marbles in urns) - There are two white marbles in urn A and four red marbles in urn B which can be interchanged. At each step of the process a marble is selected at random from each urn and the two marbles selected are interchanged.

- Find the transition matrix $Q$, and obtain its eigenvalues and its left and right eigenvectors.
- What is the probability that there are two red marbles in urn A after three steps? And after many steps?
- Express the probability vector $p^{(n)}$ in terms of the left and right eigenvectors for this problem.

### 3.2.2 Three-state chain

Consider the three state chain with transition matrix

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad (3.72)$$

The problem is to find a general formula for $Q_{11}^{(n)}$ and the invariant distribution. To find the invariant distribution we solve the vector equation $\pi = \pi Q$. This yields two scalar equations. The extra equation $\pi_1 + \pi_2 + \pi_3 = 1$ is needed to fix $\pi$ uniquely and we find $\pi = (1/5, 2/5, 2/5)$. To calculate $Q_{11}^{(n)}$ we first compute the eigenvalues of $Q$ from the characteristic equation and we get

$$\lambda_1 = 1, \quad \lambda_{2,3} = \pm \frac{i}{2} \quad (3.73)$$

Having distinct eigenvalues, $Q$ is diagonalizable, that is for some invertible matrix $U$ we get

$$Q = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{i}{2} & 0 \\ 0 & 0 & -\frac{i}{2} \end{pmatrix} U^{-1} \quad (3.74)$$

and hence

$$Q^n = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & (i/2)^n & 0 \\ 0 & 0 & (-i/2)^n \end{pmatrix} U^{-1} \quad (3.75)$$

which forces $Q_{11}^{(n)}$ to have the form

$$Q_{11}^{(n)} = a + b \left(\frac{i}{2}\right)^n + c \left(\frac{-i}{2}\right)^n \quad (3.76)$$

for some constants $a, b, c$. The answer we want is real so it makes sense to rewrite $Q_{11}^{(n)}$ in the form

$$Q_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left\{ \beta \cos \frac{n \pi}{2} + \gamma \sin \frac{n \pi}{2} \right\} \quad (3.77)$$
for constants $\alpha, \beta, \gamma$. The first few values of $Q_{11}^{(n)}$ are easy to obtain so we get equations for $\alpha, \beta, \gamma$:

$$1 = Q_{11}^{(0)} = \alpha + \beta \tag{3.78}$$

$$0 = Q_{11}^{(1)} = \alpha + \frac{1}{2} \gamma \tag{3.79}$$

$$0 = Q_{11}^{(2)} = \alpha - \frac{1}{4} \beta \tag{3.80}$$

so $\alpha = 1/5, \beta = 4/5, \gamma = -2/5$ and

$$Q_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left\{ \frac{4}{5} \cos \frac{n\pi}{2} - \frac{2}{5} \sin \frac{n\pi}{2} \right\} \tag{3.81}$$

Alternatively, we could have used the fact that $Q_{ij}^{(n)} \to \pi_j$ as $n \to \infty$, for all $i, j$, to identify $a = 1/5$, as

$$\lim_{n \to \infty} Q_{11}^{(n)} = \pi_1 = \frac{1}{5} \tag{3.82}$$

More generally, the following method may in principle be applied to find the formula for $Q_{ij}^{(n)}$ for any $M$-state chain and any state $i, j$.

1. Compute the eigenvalues $\lambda_1, ..., \lambda_M$ of $Q$ by solving the characteristic equation.

2. If the eigenvalues are distinct then $Q_{ij}^{(n)}$ has the form

$$Q_{ij}^{(n)} = a_1 \lambda_1^n + ... + a_M \lambda_M^n \tag{3.83}$$

for some constants $a_1, ..., a_M$ (depending on $i, j$). If an eigenvalue $\lambda$ is repeated (once, say) then the general form includes the term $(a + bn)\lambda^n$.

3. Complex conjugate eigenvalues are best written using sine and cosine.

**Exercise** (Random walk on a graph) -

Consider a graph of $N$ nodes, and connectivity matrix $c$. We denote the local degrees $k_i = \sum_j c_{ij}$. A random walk on this graph is a Markov chain with transition matrix

$$Q_{ij} = \begin{cases} \frac{1}{k_i} & \text{if } c_{ij} = 1 \\ 0 & \text{if } c_{ij} = 0 \end{cases} \tag{3.84}$$

(a) Justify (3.84).

(b) Show that random walks on graph are reversible Markov chains, with stationary distribution $\pi = (\pi_1, ..., \pi_N)$, $\pi_i = k_i/\sum_i k_i$.

**Exercise** A mouse lives in a house of three rooms, A, B, C. There are three doors between room A and B, two doors between room B and C, and one door between room A and C, as shown in figure (3.1).
At regular time intervals, a door, in the room occupied by the mouse, is opened at random and the mouse is trained to change room each time. After the mouse changes room the door is closed.

(a) Find the transition matrix \( Q \), i.e. the probabilities for the mouse to move from one room to another.

(b) Approximately what fraction of its time will the mouse spend in each room?

(c) If the mouse wants to spend the same fraction of its time in each room, it cannot make a transition each time a door is opened, but sometimes it will need to reject the proposed move and stay in the room it is occupying when the door is opened (see Fig. 3.2). In other words, we must introduce a probability \( A(Y|X) \) that the mouse accept the proposed move from room \( X \) to room \( Y \) (and changes room) and a probability \( 1 - A(Y|X) \) that the mouse rejects the move (and stays in \( X \)).

We call \( M : X \rightarrow Y \) the elementary move by which the mouse changes room, so \( Y = M(X) \). There is only a limited set \( D_X \) of moves \( M \) which can act on room \( X \). Let us denote \( |D_X| \) the number of moves that can act on state \( X \) (i.e. the number of doors through which the mouse can escape room \( X \)). Also, all the moves \( M \) are invertible, i.e. for each move
3.2. EXAMPLES AND EXERCISES ON MARKOV CHAINS

$M \in D_X$ there exists a unique $M^{-1}$ such that $M^{-1}[M(X)] = M[M^{-1}(X)] = X$. We can write the transition probability to go from room $X$ to room $Y$

$$Q_{XY} = \sum_{M \in D_X} p(M|X) \left[ A(M(X)|X)\delta_{Y,M(X)} + [1 - A(M(X)|X)\delta_{Y,X}] \right] \quad (3.85)$$

where $p(M|X)$ is the probability that move $M$ is proposed to the mouse, when it is in room $X$. Since each time moves are drawn at random and with equal probabilities from those that are allowed to act (each door is opened with the same probability), we have

$$p(M|X) = \frac{1}{|D_X|} \quad (3.86)$$

Working out the detailed balance condition, upon writing the equilibrium state $\pi = (1/3, 1/3, 1/3)$, leads to the following condition for the acceptance probabilities:

$$(\forall X, \forall M \in D_X) \quad p(M|X)A(M(X)|X) = p(M^{-1}|M(X))A(X|M(X)) \quad (3.87)$$

Find the transition matrix $Q$.

**Hint:** Acceptance probabilities of the form

$$A(Y|X) = \frac{|D_X|}{|D_X| + |D_Y|} \quad (3.88)$$

satisfy the relation

$$\frac{1}{|D_X|}A(Y|X) = \frac{1}{|D_Y|}A(X|Y) \quad (3.89)$$

The choice of acceptance probability is not unique. Equation (3.88) corresponds to the Glauber choice

$$A(Y|X) = \frac{P(M|Y)}{P(M|X) + P(M|Y)} \quad (3.90)$$

Another popular choice is the Metropolis one

$$A(Y|X) = \min \left\{ 1, \frac{P(M|Y)}{P(M|X)} \right\} \quad (3.91)$$
### 3.3 One-step processes

In the previous section we considered processes in discrete time. Now we turn to continuous time processes with discrete state space. The methods to be used will be similar, with transitions occurring between times \(i\) and \(i + 1\) now replaced by the ones occurring in a small time interval \((t, t + \Delta t]\).

An important family of Markov processes in continuous time and discrete state space are the generation-recombination or birth-death processes, which in short we call one-step processes. For these processes the range consists of integer \(n\) and only jumps between adjacent states are permitted. We define \(g_n = W(n + 1|n, t)\) the probability per unit time for a jump from state \(n\) to state \(n + 1\) and \(r_n = W(n - 1|n, t)\) the probability per unit time for a jump from \(n\) to \(n - 1\).

Therefore, for infinitesimal time interval \(\Delta t\), the following transition probability laws hold:

\[
\begin{align*}
P_{1|1}(n + 1, t + \Delta t|n, t) &= g_n \Delta t + O(\Delta t) \\
P_{1|1}(n - 1, t + \Delta t|n, t) &= r_n \Delta t + O(\Delta t) \\
P_{1|1}(n, t + \Delta t|n, t) &= 1 - (g_n + r_n) \Delta t + O(\Delta t) \\
P_{1|1}(n + k, t + \Delta t|n, t) &= O(\Delta t^2) \quad \forall |k| > 1 \quad (3.92)
\end{align*}
\]

We can easily write the Master equation for the evolution of the probability distribution \(p_n(t)\) that the system is in state \(n\) at time \(t\). The probability of finding the system in state \(n\) at time \(t + \Delta t\) is given by the sum of probabilities of all previous states multiplied by the probability of transition to state \(n\). For small \(\Delta t\)

\[
p_n(t + \Delta t) = [p_{n+1}r_{n+1} + p_{n-1}r_{n-1} + (1 - g_n - r_n)p_n] \Delta t \quad (3.93)
\]

Letting \(\Delta t \to 0\) we find the master equation for one-step processes

\[
\dot{p}_n = \lim_{\Delta t \to 0} \frac{p_n(t + \Delta t) - p_n(\Delta t)}{\Delta t} = r_{n+1}p_{n+1} + g_{n-1}p_{n-1} - (r_n + g_n)p_n \quad (3.94)
\]

One step processes occur at e.g. birth and death of individuals, generation and recombination processes of charge carriers, absorption or emission of photons or particles, arrival and departures of customers. The name does not imply that it is impossible for \(n\) to jump by two or more units in a time \(\Delta t\), but only that the probability for this to happen is \(O(\Delta t^2)\).

If \(n = 0\) is a boundary, equation (3.94) is meaningless for \(n = 0\) and has to be replaced with

\[
\dot{p}_0 = r_1p_1 - g_0p_0 \quad (3.95)
\]

Alternatively, it is possible to declare (3.94) valid for \(n = 0\) with the added definition \(r_0 = g_{-1} = 0\). Similarly, an upper boundary \(N\) requires

\[
\dot{p}_N = g_{N-1}p_{N-1} - r_Np_N \quad (3.96)
\]

or \(r_{N+1} = g_N = 0\).

One step processes can be subdivided in the following categories, depending on the properties of the coefficients \(r_n\) and \(g_n\).
3.3. ONE-STEP PROCESSES

- Linear, if the coefficients are linear functions of $n$
- Non-linear, if the coefficients are non-linear functions of $n$
- Random-walks, if the coefficients are constant

3.3.1 The Poisson process

The Poisson process is a random walk over $n = 0, 1, ...$ with steps to the right only, occurring at random times. It is defined by

\[ r_n = 0, \quad g_n = q, \quad p_n(0) = \delta_{n,0} \quad (3.97) \]

i.e. no recombination exists and $q$ is constant. The kronecker delta merely expresses that the probability for no events to have occurred after time zero equals one, and the probability of more than one event occurring after time zero equals zero.

The Poisson process calculates the probability of $n$ independent events occurring at time $t > 0$. These events could be for example the tunneling of electrons through a single barrier (shot noise). If we indicate by $N(t)$ the number of events occurring between the initial time $t = 0$ and time $t$, each sample function $N(t)$ takes only integer values $n = 0, 1, ...$ and is a succession of steps of unit height at random moment. These time points constitute a random set of dots on the time axis and their number between any two times $t_1$ and $t_2$ is distributed according to the Poissonian distribution. Since the events are independent, there is a probability $q\Delta t$ of a step to occur in $(t, t + \Delta t]$ regardless of what happened before, so $N(t)$ is Markov. The Master equation for Poisson process has the form

\[ \dot{p}_n = q(p_{n-1} - p_n) \quad (3.98) \]

Master equations depending on discrete stochastic variables are most easily solved through use of generating function as

\[ F(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n \quad (3.99) \]

Various moments of the stochastic variable $n$ are obtained by taking derivatives of $F(z,t)$ with respect to $z$ and allowing $z \to 1$. For example

\[ \langle n \rangle = \lim_{z \to 1} \frac{\partial F(z,t)}{\partial z} = \sum_{n=0}^{\infty} np_n(t) \quad (3.100) \]

and

\[ \langle n^2 \rangle - \langle n \rangle = \lim_{z \to 1} \frac{\partial^2 F(z,t)}{\partial z^2} = \sum_{n=0}^{\infty} (n^2 - n)p_n(t) \quad (3.101) \]

An alternative form of the last two equations is often convenient

\[ \langle n \rangle = \lim_{z \to 1} \frac{\partial \log F(z,t)}{\partial z} \]

\[ \langle n^2 \rangle - \langle n \rangle^2 - \langle n \rangle = \lim_{z \to 1} \frac{\partial^2 F(z,t)}{\partial z^2} \quad (3.102) \]
CHAPTER 3. MARKOV CHAINS AND ONE STEP PROCESSES

Multiplying the master equation by $z^n$ and summing over $n$ we get
\[
\frac{\partial F(z,t)}{\partial t} = \sum_n z^n \dot{p}_n(t) = \sum_n z^n [q(p_{n-1} - p_n)]
\]
\[
= q \sum_{n=1}^\infty z^{n-1} p_{n-1}(t) - q \sum_{n=0}^\infty z^n p_n(t)
\]
so
\[
\frac{\partial}{\partial t} F(z,t) = q(z-1)F(z,t)
\]
and the solution
\[
F(z,t) = F(z,0)e^{q(z-1)t}
\]
contains an arbitrary function $F(z,0)$ which is determined from initial conditions. If we take as our initial distribution $p_n(0) = \delta_{n,0}$ this is sufficient because, due to invariance for a shift in $n$, this also covers the case $p_n(0) = \delta_{n,m}$ for every $m$, and by suitable superposition of these, any initial distribution can be reproduced. We then get $F(z,0) = 1$ and we find $p_n(t)$ by expanding $F(z,t)$ in powers of $z$
\[
F(z,t) = q^{(z-1)t} = e^{qzt}q^{-qt} = \sum_{n=0}^\infty \frac{(qzt)^n}{n!}e^{-qt} = \sum_{n=0}^\infty z^n p_n(t)
\]
where
\[
p_n(t) = \frac{(qt)^n}{n!}e^{-qt}
\]

We now show that the time intervals between successive events (times at which unit steps occur) are (independently) distributed with the exponential probability distribution function $qe^{-qt}$. Set $t_0$ as the new time origin. An event may or may not have occurred at $t_0$. Any property of the process after $t_0$ is independent of what happened at $t_0$ or before. Let $t_0 + t$ be the time at which the first event occur after $t_0$. Calculate the probability distribution of $t$. Define $P(T) = P(t > T)$.
\[
P(T + \Delta T) = P(t > T + \Delta T) = P(t > T \text{ and no event occurs in } (t_0 + T, t_0 + T + \Delta T))
\]
\[
= P(t > T)P(\text{no event occurs in } (t_0 + T, t_0 + T + \Delta T)|t > T)
\]
\[
= P(t > T)P(\text{no event occurs in } (t_0 + T, t_0 + T + \Delta T))
\]
\[
= P(T)(1 - q\Delta t + O(\Delta T))
\]
or
\[
P'(T) = -qP(T) \rightarrow P(T) = e^{-qT}
\]
where we used
\[
P(0) = P(t > 0) = 1
\]
So
\[
P(t) = P(T > t) = 1 - P(T) = 1 - e^{-qT}
\]
The probability density $W_T(t)$ follows as
\[
W_T(t) = \frac{dP}{dt} = qe^{-qT}
\]
Because $\langle t \rangle = q^{-1}$ it follows that the transition probability in the master equation equals the inverse expectation time for the occurrence of an event.
3.4 Examples and exercises on One step processes

3.4.1 The decay process

An example of a linear one-step process is the decay process.

Exercise Consider a piece of radioactive material. The number of active nuclei surviving at time \( t > 0 \), \( N(t) \), is a non-stationary Markov process. Find the time evolution for the average number of survived nuclei.

Answer

Let \( p_n(t) \) be the probability that there are \( n \) surviving nuclei at time \( t \). If \( \gamma \) is the decay probability per unit time for one nucleus, the transition probability from \( m \) to \( n \) in a short time \( \Delta t \) is

\[
P_{1|1}(n, t + \Delta t|m, t) = \delta_{m,n}(1 - \gamma m\Delta t) + \delta_{m-1,n}\gamma m\Delta t + \mathcal{O}(\Delta t^2)
\]

which leads to the master equation

\[
\dot{p}_n(t) = \gamma(n + 1)p_{n+1}(t) - \gamma np_n(t)
\]

We now apply a device for linear Master equation which consists in multiplying both sides of (3.114) by \( n \) and summing over \( n \), thus obtaining

\[
\sum_{n=0}^{\infty} np_n(t) = \gamma \sum_{n=0}^{\infty} n(n + 1)p_{n+1} - \gamma \sum_{n=0}^{\infty} n^2 p_n = \gamma \sum_{n=0}^{\infty} n - \gamma \sum_{n=0}^{\infty} n^2 p_n = -\gamma \sum_{n=0}^{\infty} n p_n
\]

Thus we have found the mean field equation for \( N(t) \)

\[
\frac{d\langle N(t) \rangle}{dt} = -\gamma \langle N(t) \rangle
\]

Solving the above equation for \( \langle N(0) \rangle = n_0 \) gives

\[
\langle N(t) \rangle = n_0 e^{-\gamma t}
\]

3.4.2 Birth-death process

Linear and non-linear birth-death processes occur very commonly in chemistry and population dynamics. Let us consider a population of \( m \) bacteria at time \( t \) such that

- The probability of a bacterium dying in time \( t \to t + \Delta t \) is \( r_m \Delta t \)
- The probability of a bacterium being created in time \( t \to t + \Delta t \) is \( g_m \Delta t \)
- The probability of no change in the number of bacteria in time \( t \to t + \Delta t \) is \( 1 - (g_m + r_m) \Delta t \)
- The probability of more than one birth or death in time \( t \to t + \Delta t \) is zero.
Exercise Obtain a partial differential equation for the generating function.

Answer
We can write
\[ P_{1|1}(n, t + \Delta t|m, t) = (1 - (g_m + r_m)\Delta t)\delta_{n,m} + (g_m\delta_{n,m+1} + r_n\delta_{n,m-1})\Delta t \] (3.118)

Let us now assume that the probability of a birth or death is proportional to the number of bacteria present so that \( g_m = mg \) and \( r_m = mr \). We obtain the master equation for the linear birth-death process
\[ \frac{\partial P_1(n, t)}{\partial t} = (n - 1)gP_1(n - 1, t) + (n + 1)rP_1(n + 1, t) - (ng + nr)P_1(n, t) \] (3.119)

It is linear because the coefficient on the right hand side are linear in \( n \). Solution can again be found by using generating function
\[ F(z, t) = \sum_{n=-\infty}^{\infty} p_n(t)z^n \] (3.120)

In terms of \( F \) the master equation reads
\[ \frac{\partial F}{\partial t} = (z - 1)(gz - r)\frac{\partial F}{\partial z} \] (3.121)

which is a first-order linear partial differential equation and may be solved by using the methods of characteristics.

3.4.3 Chemical reaction
An example of nonlinear one step process is a chemical reaction of the form
\[ X \rightarrow 2X \] with rate \( k \) (3.122)
\[ X \leftarrow 2X \] with rate \( k' \) (3.123)

In this case, for state \( n \), the generation probability per unit time is \( g_n = kn \). For the reverse reaction, one can assemble pairs of molecules of \( X \) in \( n(n - 1) \) ways, thus \( r_n = n(n - 1)k' \).

Exercise Show that the Master equation is
\[ p_n = k'n(n + 1)p_{n+1} + k(n - 1)p_{n-1} - knp_n - k'n(n - 1)p_n \] (3.124)

and show that the equation for the average \( \langle n \rangle \) is not closed (due to nonlinearity).

3.4.4 Continuous time random walk
Here we present a continuous time formulation of the random walk, which can be described by a master equation. A walker takes steps left or right with a probability per unit time \( q \), which means that the walker waits at each point for a variable time.
3.4. EXAMPLES AND EXERCISES ON ONE STEP PROCESSES

For unbounded symmetrical random walk \( r_n = g_n = q \). The transition from \( m \) to \( n \) in a short time \( \Delta t \) is

\[
P(n, t + \Delta t|m, t) = q \Delta t \delta_{n,m+1} + q \Delta t \delta_{n,m-1} + (1 - 2q \Delta t) \delta_{n,m}
\]

and master equation results

\[
\frac{\partial p_n(t)}{\partial t} = qp_{n-1}(t) + qp_{n+1}(t) - 2qp_n(t)
\]

The constant \( q \) can be absorbed into the time unit. This simple example often suffices to illustrate more complicated processes. In particular it contains the essential features of a physical diffusion process.

**Exercise** Show, by using generating function, that the solution of (3.126) is

\[
p_n(t) = e^{-2tq} \sum_{l=0}^{\infty} \frac{t^{2l+n}}{(l+n)!} = e^{-2tq} I_n(2tq)
\]

where \( I_n \) is the \( n \)-th modified Bessel function. Verify that \( \langle n(t) \rangle = 0 \) and \( \langle n^2(t) \rangle = 2t \), typical of diffusion processes.

### 3.4.5 Diffusion-annihilation on complex networks

Consider the diffusion-annihilation process \( A + A \rightarrow 0 \) on a complex network of size \( N \), fully defined by the adjacency matrix \( c_{ij} \), which takes the values \( c_{ij} = 1 \) if vertices \( i \) and \( j \) are connected by an edge, and 0 otherwise. The network is undirected, so \( c_{ij} = c_{ji} \). We denote

\[
k_i = \sum_j c_{ij} \quad \text{the degree of node } i
\]

\[
p(k) = \sum_i \delta_{k,k_i} \quad \text{the degree distribution}
\]

\[
\langle k \rangle = \sum_k kp(k) \quad \text{the average degree}
\]

\[
W(k, k') = \frac{\sum_{ij} c_{ij} \delta_{k,k_i} \delta_{k',k_j}}{N(k)} \quad \text{the degree correlation}
\]

Each vertex in the network can host at most one \( A \) particle and the dynamics of the process is defined as follows: each particle jumps at a certain rate \( \lambda \) to a randomly chosen nearest neighbour. If the latter is empty the particle fills it, leaving the first vertex empty. If the nearest neighbour is occupied the two particles annihilate, leaving both vertex empty.

**Exercise** Let \( n_i(t) \) be a binary random variable taking values 0 if vertex \( i \) is empty and 1 if vertex \( i \) is occupied. The state at time \( t \) is completely defined by the state vector \( n(t) = (n_1(t), n_2(t), ..., n_N(t)) \).

(a) Assuming that the time evolution of the particles follows a Poisson process, show that the evolution of \( n_i(t) \) after a time increment \( dt \) can be expressed as

\[
n_i(t + dt) = n_i(t)\theta(dt) + [1 - n_i(t)]\sigma(dt)
\]
where $\theta(dt)$ and $\sigma(dt)$ are binary random variables taking values

$$
\theta(dt) = \begin{cases} 
0 & \text{with probability } \lambda dt \left[ 1 + \sum_j \frac{c_{nj}(t)}{k_j} \right] \\
1 & \text{otherwise}
\end{cases}
$$

and

$$
\sigma(dt) = \begin{cases} 
1 & \text{with probability } \lambda dt \sum_j \frac{c_{nj}(t)}{k_j} \\
0 & \text{otherwise}
\end{cases}
$$

(b) Find an equation for $\langle n_i(t + dt) \rangle|_{n(t)}$, i.e. for the average evolution of the system, conditioned to the knowledge of its state at the previous time step. By taking the average over all the possible configurations $n$, derive the following master equation for the density $\rho_i(t) = \langle n_i(t) \rangle$ of A particle at site $i$:

$$
\frac{d\rho_i(t)}{dt} = -\rho_i(t) + \sum_j c_{ij} \frac{1}{k_j} [\rho_j(t) - 2\rho_{ij}(t)]
$$

where $\rho_{ij}(t) = \langle n_i(t)n_j(t) \rangle$.

(c) Assume now that all the vertices with the same degrees are statistically equivalent, i.e.

$$
\rho_i(t) \equiv \rho_k(t) \forall i \text{ such that } \sum_\ell c_{i\ell} = k
$$

$$
\rho_{ij}(t) \equiv \rho_{kk'}(t) \forall i, j \text{ such that } \sum_\ell c_{i\ell} = k, \text{ and } \sum_\ell c_{j\ell} = k'
$$

Show that the density $\rho_k(t)$ of A particles on nodes of degree $k$ evolves according to

$$
\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \frac{\langle k \rangle}{p(k)} \sum_{k'} \frac{\rho_{kk'}(t)}{k'} W(k, k')
$$
Chapter 4

Topics in equilibrium and nonequilibrium statistical mechanics

Here we set up the machinery for a microscopic probabilistic description of matter, i.e. lay the foundations of statistical mechanics. We shall consider classical systems with a large number of degrees of freedom such as $N$ interacting particles in a box or $N$ interacting objects in a lattice. The motion of such objects is described by Newton’s laws or, equivalently, by Hamiltonian dynamics. In 3 dimensions, such a system is specified by $6N$ independent positions and momentum coordinates. In the $6N$-dimensional phase space the state of the system is given by a single point which moves according to Hamiltonian dynamics, as the state of the system changes. If we are given a real $N$-particle system, we never know exactly what its state is. We only know with a certain probability that it is one of the points in the phase space. Thus, the state point can be regarded as a stochastic variable and we can assign a probability distribution to the points in phase space in accordance with our knowledge of the state of the system.

We can then view the phase space as a probability fluid which flows according to Hamiltonian dynamics. In this way we obtain a connection between the Mechanical description and a Probabilistic description, and the problem of finding an equation of motion for the probability density reduces to a problem in fluid mechanics. For Hamiltonian systems, the phase space probability distribution evolves according to Liouville equation.

The $N$-body probability density for a classical system contains much more information about the system than we would ever need or want. In practice, the main use of the probability density is to find expectation values or correlation functions for various observables, since those are what we measure experimentally and what we deal with in thermodynamics. As a consequence of the discrete nature of matter, the system receives random “kicks” from density fluctuations continually, even when it is in thermodynamic equilibrium, so systems in equilibrium undergo fluctuations about their equilibrium state. Information about these fluctuations is contained in the dynamic correlation function for the equilibrium system, and in the power spectrum of the equilibrium fluctuations. The power spectrum is a quantity often measured in experiments. The time reversal invariance of the underlying Newtonian dynamics of the various degrees of freedom puts constraints on the behaviour of the dynamic equilibrium fluctuations and the time-dependent correlation functions that characterize these fluctuations.
Systems which are out of equilibrium generally return to the equilibrium state through a variety of processes which may or may not be coupled to one another. Generally there are a few degrees of freedom that decay back to equilibrium very slowly. Fluctuations about the equilibrium state decay on the average according to the same macroscopic laws that govern the decay of a non-equilibrium system to the equilibrium state. If we can probe equilibrium fluctuations we have a means of probing the decay process in a system. Linear response theory provides a tool for probing equilibrium fluctuations by applying a weak external field which couples to the system. The system responds to the field in a manner that depends entirely on the spectrum of equilibrium fluctuations. The response to the dynamic field is measured by the susceptibility. The fluctuation-dissipation theorem links the susceptibility to the correlation for equilibrium fluctuations. The spectrum of equilibrium fluctuations determines the rate of absorption of energy from the external field.

4.1 Equilibrium

Consider an isolated system whose energy is specified to lie in a narrow range. Let \( \Omega \) be the number of states accessible to the system. In equilibrium this system is equally likely to be found in any of these states (postulate of equal a priori probability). If a removal of constraints result in increasing the number of states accessible to the system \( \Omega_f > \Omega_i \), immediately after the constraints are removed, the system occupies only a fraction \( P = \Omega_i / \Omega_f \) of the \( \Omega_f \) states now accessible. This is not an equilibrium situation. If \( \Omega_f >> \Omega_i \), then there is a pronounced tendency for the situation to change in time until the system in the ensemble are distributed equally over all the possible \( \Omega_f \) states (and equilibrium distribution of systems over accessible states is attained). In this case equilibrium does not prevail at all stage of the process and the process is irreversible (in this case the removal of constraints cannot restore the initial situation). Irreversible thermodynamics comprises the phenomenologic study of the approach to equilibrium and the behaviour of, and conditions for, a steady state in a system close to equilibrium.

4.2 Probability distributions in dynamical systems

The state of a closed classical system with \( 3N \) degrees of freedom (for example, \( N \) particles in a three-dimensional box) is specified by a set of \( 6N \) independent real variables \( X^N = X^N(q^1, ..., q^N, p^1, ..., p^N) \), where \( q_j \) and \( p_j \) are the position and momentum of the \( j \)th particle. The time evolution of \( (q^i, p^i) \) is governed by the Hamiltonian \( H(X^N, t) \)

\[
\dot{q}^i = \frac{\partial H}{\partial p^i} \\
\dot{p}^i = -\frac{\partial H}{\partial q^i}
\]  

(4.1)

If the state \( X^N \) is known at time \( t \), it is completely determined for any other time (the motion is deterministic so it is uniquely determined from initial conditions). If the Hamiltonian does not depend explicitly on the time, i.e. \( H = H(X^N) \), the system is conservative and the energy is conserved. Let \( \Gamma \) be the \( 6N \)-dimensional phase space. The state vector \( X^N \) is a point in \( \Gamma \).
As the state evolves in time, \( X^N \) traces out a trajectory in \( \Gamma \) (different trajectories do not cross because the system is deterministic.)

When we deal with real systems we can never specify the state of the system. There is always some uncertainty in the initial conditions. So we consider \( X^N \) as a stochastic variable and introduce a probability density \( \rho(X^N, t) \) on the phase space, where \( \rho(X^N, t) \) is the probability that the state point \( X^N \) lies in the volume element \( X^N, X^N + dX^N \) at time \( t \). We introduce a picture of \( \Gamma \) filled with a fluid of state points. Each point of fluid is assigned a probability in accordance with our knowledge of the system and carry this probability for all time (probability is conserved). The change in our knowledge corresponds to the fluid flowing. State points must always lie somewhere so

\[
\int dX^N \rho(X^N, t) = 1 \quad (4.2)
\]

If at some time there is only a small uncertainty in the state of the system, the probability density will be sharply peaked in the region where the state is known to be located and zero elsewhere. As time passes either probability remains sharply peaked (although the peak region can move) and we do not lose knowledge about the state of the system, or it spreads out and becomes rather uniformly distributed (all knowledge is lost). The probability behaves like a fluid in phase space. So we shall use fluid mechanics to obtain equations for probability densities.

Consider a small volume \( V_0 \) at a fixed location in \( \Gamma \). Since the probability is conserved, the total decrease in probability in \( V_0 \) per unit time is entirely due to the flow of probability through the surface of \( V_0 \)

\[
\frac{dP(V_0)}{dt} = \frac{\partial}{\partial t} \int_{V_0} dX^N \rho(X^N, t) = -\oint_S dS^N \rho(X^N, t) \dot{X}^N \quad (4.3)
\]

By using Gauss theorem we get

\[
\frac{\partial}{\partial t} \int_{V_0} dX^N \rho(X^N, t) = -\int_{V_0} dX^N \nabla_{X^N} (\rho(X^N, t) \dot{X}^N) \quad (4.4)
\]

with

\[
\nabla_{X^N} = \left( \frac{\partial}{\partial q^1}, \ldots, \frac{\partial}{\partial q^N}, \frac{\partial}{\partial p^1}, \ldots, \frac{\partial}{\partial p^N} \right) \quad (4.5)
\]

Thus we obtain the *continuity equation*

\[
\frac{\partial}{\partial t} \rho(X^N, t) + \nabla_{X^N} (\rho(X^N, t) \dot{X}^N) = 0 \quad (4.6)
\]

that, we rewrite

\[
\frac{\partial \rho}{\partial t} + \sum_i \left( \frac{\partial}{\partial q^i} (\rho q^i) + \frac{\partial}{\partial p^i} (\rho p^i) \right) = 0 \quad (4.7)
\]

The second term on the LHS evaluates to

\[
\dot{q}^i \frac{\partial \rho}{\partial q^i} + \dot{p}^i \frac{\partial \rho}{\partial p^i} \quad (4.8)
\]
because
\[ \rho \left( \frac{\partial \dot{q}^i}{\partial q^i} + \frac{\partial \dot{p}^i}{\partial p^i} \right) = \rho \left( \frac{\partial^2 H}{\partial q^i \partial \dot{p}^i} - \frac{\partial^2 H}{\partial p^i \partial \dot{q}^i} \right) = 0 \] (4.9)
so finally
\[ \frac{\partial \rho}{\partial t} + \sum_i \left( \dot{q}^i \frac{\partial \rho}{\partial q^i} + \dot{p}^i \frac{\partial \rho}{\partial p^i} \right) = 0 \] (4.10)
Note that this equation gives the time rate of change of \( \rho(X^N, t) \) at a fixed point in phase space.
We can rewrite the continuity equation in terms of the operator
\[ L = \sum_i \left( \frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \right) \] (4.11)
as
\[ i \frac{\partial \rho}{\partial t} = -L \rho \] (4.12)
or, by introducing the Liouville operator \( L = -iL \), as
\[ i \frac{\partial \rho}{\partial t} = L \rho \] (4.13)
which is called Liouville equation. The Liouville operator is a Hermitian differential operator. If we know the probability distribution at time \( t = 0 \) we can know the probability density at time \( t \)
\[ \rho(X^N, t) = e^{-iLt} \rho(X^N, 0) \] (4.14)
A probability density which remains constant (stationary) in time satisfies \( L \rho_{\text{stat}} = 0 \). Liouville equation implies that the volume of a region in the phase space which evolves according to Hamilton equation is conserved. Following Gibbs, the probability density is often interpreted in terms of the concept of ensemble of systems. If we look at each system at a given time, it will be represented by a point in the \( 6N \)-dimensional phase space. The density of points representing an ensemble of \( n \) systems in phase space is given by \( n \rho(X^N, t) \). Equation (4.14) gives the general equation for the evolution of the probability density of a classical dynamical system. However it is different in an important respect from the Master equation we have derived in the pervious chapter. Liouville operator is Hermitian so has real eigenvalues. So the solutions (4.14) will oscillate and not decay to a unique equilibrium state. Furthermore, if we reverse the time we do not change the equations of motion for the probability density, since the Liouville operator \( L \) changes the sign under time reversal. This is different from the Master equation and the corresponding Fokker-Planck equation which will change into a new equation under time reversal. Equation (4.14) does not admit an irreversible decay of the system to a unique equilibrium state, as we commonly observe in nature. The problem of obtaining irreversible decay to equilibrium from the Liouville equation, is one of the central problem of statistical physics and it is beyond the scope of this course.

### 4.3 Detailed balance

In this section we prove detailed balance for closed, isolated, classical systems, i.e. for systems which can be described in terms of an Hamiltonian, for which energy is a constant of motion
(so that the trajectories in \( \Gamma \) are confined to single energy shells) and for which no exchange of matter takes place (so that the set of microscopic variables is fixed).

The law of microscopic reversibility states that in equilibrium a process and its reverse will occur with the same frequency. Consider a 2-dimensional system with \( N \) degrees of freedoms. A microstate (configuration) will be represented by a point in the \( 2N \)-dimensional phase space. Let \( Y(q,p) \) be a set of macroscopic observables. Let \( A \) be a set of microscopic configurations (phase points) such that \( Y(q,p) = y \). Assume \( Y \) are even functions of the moments \( Y_{X} = Y_{\bar{X}} \).

In discrete state space

\[
W_{nn'}p_{n'} = W_{n'n}p_{n} \quad \forall \ n, n'
\]  

Detailed balance holds for closed, isolated, classical systems in equilibrium with

- Hamiltonian is an even function of all moments (If \( H \) is not even in the moments, e.g. because of a field, we could proceed as before, but we would have to revert the field as well.)

- Observable \( Y \) is an even function of all moments.

For systems which are not in equilibrium we cannot apply the laws of microscopic reversibility and detailed balance does not work. However, a system can be brought to a steady state by outside influences (gradients maintained by reservoirs). In this case equation

\[
\sum_{n'} W_{nn'}p_{n'} = \sum_{n'} W_{n'n}p_{n} 
\]  

holds and we get in the master equation

\[
\frac{dp_{n}}{dt} = \sum_{n'} (W_{nn'}p_{n'} - W_{n'n}p_{n}) = 0
\]
4.4 Time dependent correlations

Microscopic reversibility implies that the time dependent correlation function of mesoscopic fluctuations \( x = (x_1, \ldots, x_N) \), \( x_i = X_i - X_i^0 \), of the state variables \( X_i \), about the equilibrium value \( X_i^0 \), obey the relation

\[
\langle x_i(0)x_j(t) \rangle = \langle x_i(t)x_j(0) \rangle
\]  

In fact, we have for the correlation matrix

\[
C(t) = \langle x(0)x(t) \rangle = \int dx\, dx'\, xx' P(x, 0) P(x', t|x, 0) = \int dx\, dx'\, xx' P(x', 0) P(x, t|x', 0) = \langle x(t)x(0) \rangle
\]  

so \( C(t) = C^T(t) \). For systems governed by a stationary distribution (such as systems in equilibrium) we also have

\[
C(t) = \langle x(t)x(0) \rangle = \langle x(t+T)x(T) \rangle = \langle x(0)x(-t) \rangle = \langle x(-t)x(0) \rangle = C^T(-t)
\]  

where in the last line we set \( T = -t \). Microscopic reversibility and stationarity therefore imply

\[
C(t) = C(-t)
\]

4.5 Wiener-Khintchine theorem

The Wiener-Khintchine theorem enables us to obtain a relation between the correlation matrix of time dependent fluctuations of a certain quantity \( X(t) \) and its spectral density matrix for ergodic systems with stationary distribution functions. The spectral density matrix of a fluctuating quantity \( X(t) \) is defined as

\[
S(\omega) = \lim_{T \to \infty} \frac{1}{T} |\bar{X}(\omega; T)|^2 \quad \text{where} \quad \bar{X}(\omega; T) = \int_{-T}^{T} dt \, e^{-i\omega t} X(t; T)
\]  

and

\[
X(t; T) = \begin{cases} 
X(t) & |t| < T \\
0 & |t| \geq T
\end{cases}
\]

Because \( X(t) \) is real,

\[
|\bar{X}(\omega; T)|^2 = \bar{X}^*(\omega; T)\bar{X}(\omega; T) = \bar{X}(-\omega; T)\bar{X}(\omega; T)
\]

so

\[
S(\omega) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} dt \, e^{-i\omega t} x(t; T) \int_{-T}^{T} dt' e^{i\omega t'} x(t'; T)
\]

\[
= \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} dt \int_{-T}^{T} dt' \, e^{i\omega t} x(t; T) x(t + \tau; T) \\
= \lim_{T \to \infty} \frac{1}{T} \int_{-\infty}^{\infty} d\tau \, e^{i\omega \tau} \int_{-\infty}^{\infty} dt \, x(t; T) x(t + \tau; T)
\]  

(4.27)
If we now invoke the ergodic theorem, we can equate the time average to the ensemble average and thus express the correlation function of the fluctuating quantity $X(t)$ as

$$C(\tau) = \langle X(t)X(t+\tau) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} dt \ X(t)X(t+\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{-\infty}^{\infty} dt \ X(t;T)X(t+\tau;T)$$

(4.28)

We can then relate the spectral density matrix to the correlation matrix (Wiener-Khintchine theorem)

$$S(\omega) = \int_{-\infty}^{\infty} d\tau \ e^{i\omega \tau} C(\tau)$$

(4.29)

From time reversal invariance $C(\tau) = C(-\tau)$,

$$S(\omega) = \int d\tau \ cos(\omega \tau) C(\tau) = S(-\omega)$$

(4.30)

We conclude that the spectral density matrix $S(\omega)$ is a real and even function of $\omega$. This implies that

$$C(\tau) = \frac{1}{2\pi} \int d\omega \ e^{-i\omega \tau} S(\omega) = \frac{1}{2\pi} \int d\omega \ cos(\omega \tau) S(\omega)$$

(4.31)

### 4.5.1 White noise

The term white noise is applied, in analogy with the case of light, to any fluctuating quantity $X$ with $\delta$-correlation

$$\langle X(t)X(t') \rangle = \delta(t - t')$$

(4.32)

because the spectral distribution is then independent on the frequency $\omega$

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\tau) e^{-i\omega \tau} d\tau = \frac{1}{2\pi}$$

In the case of light, the frequencies correspond to different colors of light and if we perceive light to be white, it is found that in practice all colors are present in equal proportions (the optical spectrum of white light is thus flat) at least within the visible range. If the spectral density depends on the frequency, one uses the term colored noise.

White noise cannot actually exist (one simple demonstration is the UV catastrophe). However, models with almost flat $S(\omega)$ are very useful and much used. A typical one is

$$S(\omega) = \frac{1}{\pi(\omega^2 \tau_C^2 + 1)}$$

that is flat provided $\omega \ll \tau_C^{-1}$. The Fourier transform can be explicitly evaluated in this case (by residues) to give

$$C(\tau) = \frac{1}{\tau_C} e^{-\tau/\tau_C}$$

so that the autocorrelation function vanishes for $\tau \gg \tau_C$, which is called the correlation time of the fluctuating quantity $X$. Thus, the delta function correlation appears as an idealization, only valid on sufficiently long time scale.
CHAPTER 4. TOPICS IN EQUILIBRIUM AND NONEQUILIBRIUM STATISTICAL MECHANICS

4.6 Fluctuation-Dissipation theorem

In the next few sections we examine the relaxation toward equilibrium of a system that has been brought out of equilibrium by an external fluctuation. The main result is that for small deviation from equilibrium this relaxation is described by equilibrium time correlation function (Onsager regression law), i.e. the response of a system in thermodynamic equilibrium to a small applied force is the same as its response to a spontaneous fluctuation. So it is possible to probe the equilibrium fluctuation by applying a weak external field which couples to particles in the medium but yet is too weak to affect the medium. The system will respond to the field and absorb energy from it in a manner which depends entirely on the spectrum of the equilibrium fluctuations. The fluctuation dissipation theorem connects the irreversible dissipation of energy into heat to reversible thermal fluctuations at thermodynamic equilibrium.

It was Einstein to observe, in 1905, in his paper on Brownian motion, that the same random forces that cause the erratic motion of a particle in Brownian motion, could also cause drag if the particle were pulled through the fluid. In other words the fluctuations of the particle at rest have the same origin as the dissipative fluctuation force one must do work against if one tries to perturb the system in a particular direction.

4.6.1 Linear response

In this section we study the small deviations from equilibrium driven by a small perturbation applied to the system when the equilibrium situation is described by an unperturbed classical Hamiltonian $H_0(q,p)$. $(q,p)$ is the short-hand notation for the full set of canonical variables $(q_1, ..., q_N, p_1, ..., p_N)$, $N$ being the number of the degrees of freedom. The equilibrium average of a classical dynamical variable $A(q,p)$ is

$$\langle A \rangle_{\text{eq}} = \int dq \, dp \, A(q,p) \, P_{\text{eq}}$$

where the probability density $P_{\text{eq}}$ is the normalized Boltzmann weight

$$P_{\text{eq}} = \frac{e^{-\beta H_0(q,p)}}{Z_0}, \quad \text{with} \quad Z_0 = \int dq \, dp \, e^{-\beta H_0} \quad (4.34)$$

Let us perturb the Hamiltonian $H_0$ by applying a constant perturbation coupled to the observable $B$, such that the energy changes from $H_0$ into $H_1 = H_0 + \delta H$, with $\delta H = -B f$. $f$ is often called (external) field (or force, from forced harmonic oscillators where perturbations are in the form $\delta H = -f x$). Let $\langle \cdots \rangle$ denote average computed with the perturbed Hamiltonian. We define $\langle \delta A \rangle$ the variation of the response $\langle A \rangle$ of the system

$$\langle \delta A \rangle = \langle A \rangle - \langle A \rangle_{\text{eq}} = \frac{\int dq \, dp \, A e^{-\beta (H_0 + \delta H)}}{\int dq \, dp \, e^{-\beta (H_0 + \delta H)}} - \frac{\int dq \, dp \, A e^{-\beta H_0}}{\int dq \, dp \, e^{-\beta H_0}} \quad (4.35)$$

to a variation of the forces $f$. Since we are interested in small deviations from equilibrium Hamiltonian $H_0$, i.e. $\delta H$ is small, we can restrict to the limit $f \to 0$ and we can expand $e^{-\beta \delta H} = 1 - \beta \delta H + ... = 1 - \beta B f + ...$. If we truncate the expansion to linear orders in the
forces \( f \), we get the linear response

\[
\langle \delta A \rangle = \int dq dp A e^{-\beta H_0(q,p)} \left( 1 - \beta \frac{\int dq dp \delta H e^{-\beta H_0}}{Z_0} \right) - \langle A \rangle_{eq}
\]

\[
= -\beta \langle A \delta H \rangle_{eq} + \beta \langle A \rangle_{eq} \langle \delta H \rangle_{eq}
\]

\[
= \beta f \langle AB \rangle_{eq} - \beta f \langle A \rangle_{eq} \langle B \rangle_{eq}
\]

\[
= \beta f \langle \langle AB \rangle \rangle_{eq}
\]

(4.36)

The deviation from equilibrium is linearly related to perturbation. So far we considered \( H_1 \) time independent. Now assume that the perturbation \( f \) has been on for infinite time during which the system was in equilibrium with perturbed Hamiltonian, i.e. with the equilibrium probability distribution

\[
P_1 = \frac{e^{-\beta H_1(q,p)}}{\int dq dp e^{-\beta H_1(q,p)}}
\]

(4.37)

and is switched off at \( t = 0 \) (this introduces an explicit time dependence). We are interested in measurements in the system for \( t > 0 \) as it relaxes to equilibrium. Let

\[
H = \begin{cases} 
H_0 + \delta H \equiv H_1 & t < 0 \\
H_0 & t \geq 0
\end{cases}
\]

(4.38)

The expectation value of \( A(t) \) calculated over the nonequilibrium ensemble is given by the average over all possible dynamical paths originating from initial configurations at \( t = 0 \), weighted with their equilibrium distribution \( P_1 \). It is convenient to rewrite \( A(t) = A(t;q,p) \), where \( q = q(0) \) and \( p = p(0) \), as a function of the initial conditions. The change in the measured quantity, for \( \beta f \ll 1 \)

\[
\langle \delta A(t) \rangle = \frac{\int dq dp A(t;q,p)e^{-\beta H_0(q,p)+\beta B(0)f}}{\int dq dp e^{-\beta H_0(q,p)+\beta B(0)f}} - \frac{\int dq dp A(t;q,p)e^{-\beta H_0(q,p)}}{\int dq dp e^{\beta H_0(q,p)}}
\]

\[
= \beta f \langle \langle A(t)B(0) \rangle \rangle_{eq} - \langle A(t) \rangle_{eq} \langle B(0) \rangle_{eq}
\]

\[
= \beta f C_{AB}(t)
\]

(4.39)

where we have defined the Kubo function

\[
C_{AB}(t) = \langle \langle A(t)B(0) \rangle \rangle_{eq}.
\]

(4.40)

Note that \( C_{AB}(t) = C_{BA}(-t) \) because in absence of force the system is invariant under time shifts \( \langle \langle A(t)B(0) \rangle \rangle_{eq} = \langle \langle A(0)B(-t) \rangle \rangle_{eq} \). Equation (4.39) is normally referred to as Onsager regression law.

For general time dependent perturbing forces \( f \) we write the linear response as

\[
\langle \delta A(t) \rangle = \int_{-\infty}^{t} dt' R_{AB}(t,t')f(t')
\]

(4.41)

where we have introduced the retarded Green function

\[
R_{AB}(t,t') = \frac{\delta \langle A(t) \rangle}{\delta f(t')}
\]

(4.42)
which gives the impulse response function, i.e. the response of $\langle A(t) \rangle$ at time $t$ to a small impulse at time $t'$ in the field $f(t) = f \delta(t - t')$ conjugate to $B$.

Rather often results are presented in the frequency domain. One defines the Fourier transform and its inverse as

\[
\tilde{A}(\omega) = \int dt e^{i\omega t} A(t), \quad A(t) = \int \frac{d\omega}{2\pi} e^{-i\omega t} \tilde{A}(\omega) \tag{4.43}
\]

The linear susceptibility $\tilde{\chi}_{AB}$ is the Fourier transform of (4.42).

Function $R_{AB}$ and its Fourier transform have the following properties:

- Because of the stationarity of the unperturbed system
  \[
  R_{AB}(t, t') = R_{AB}(t - t') \tag{4.44}
  \]

- Since the response must be causal (cannot preceed the force that causes it, i.e. perturbations cannot propagate backwards in time)
  \[
  R_{AB}(t - t') = 0 \quad \text{for} \quad t < t' \tag{4.45}
  \]

- Finally, the response is real, so
  \[
  \tilde{\chi}_{AB}(\omega) = \tilde{\chi}_{AB}^*(-\omega) \tag{4.46}
  \]

Thus, if we introduce the real and imaginary part of the Fourier component of the response

\[
\tilde{\chi}_{AB}(\omega) = \tilde{\chi}_{AB}'(\omega) + i\tilde{\chi}_{AB}''(\omega), \quad \tilde{\chi}_{AB}'(\omega) = \text{Re}\tilde{\chi}_{AB}(\omega) \quad \tilde{\chi}_{AB}''(\omega) = \text{Im}\tilde{\chi}_{AB}(\omega) \tag{4.47}
\]

we immediately get

\[
\tilde{\chi}_{AB}'(\omega) = \tilde{\chi}_{AB}'(-\omega) \quad \tilde{\chi}_{AB}''(\omega) = -\tilde{\chi}_{AB}''(-\omega) \tag{4.48}
\]

i.e. the real part of the susceptibility is an even function of $\omega$, while the imaginary part is odd.

One can find relations between the real and the imaginary part of the susceptibility, Kramers-Kronig relations, which are a direct consequence of the response causality.

Using the Fourier representation, the convolution (4.41) is transformed into a product

\[
\langle \delta A(\omega) \rangle = R_{AB}(\omega)f(\omega) \tag{4.49}
\]

and we find that a force of a given frequency can only excite a response of the same frequency. This will not be true if the response depends on the force (nonlinear response).

For the constant external perturbation switched off at $t = 0$, (4.41) specializes to $f(t) = f\theta(-t)$

\[
\langle \delta A(t) \rangle = f \int_{-\infty}^{0} dt' R_{AB}(t - t') = f \int_{t}^{\infty} d\tau R_{AB}(\tau) \tag{4.50}
\]
Differentiating (4.50) with respect to time and using (4.39) then gives
\[ R_{AB}(t) = \begin{cases} -\beta \frac{d}{dt} C_{AB}(t) & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \] (4.51)
or
\[ R_{AB}(t) = -\beta \frac{d}{dt} C_{AB}(t) \theta(t) \] (4.52)

4.6.2 The fluctuation-dissipation theorem

Now we restrict to work with single dynamical variable \( A \), and we write simply \( R \) for \( R_{AA} \), so
\[ R(t) = -\beta \theta(t) \dot{C}(t) \] (4.53)

The fluctuation-dissipation theorem (FDT) provides a relation between the response function \( \chi(t) \) and the equilibrium autocorrelation function
\[ \frac{\partial \chi}{\partial t} = R(t) = -\frac{1}{T} \frac{\partial C(t)}{\partial t} \] (4.54)
(Boltzmann constant has been here set to one).

Often, the FDT is written in terms of two-time dynamical quantities \( R(t, t') \) and \( C(t, t') = \langle A(t)A(t') \rangle \), where \( t' \) is normally thought of as a waiting time from system preparation. In equilibrium, due to time translation invariance we have \( C(t, t') = C(t-t') \) and \( R(t, t') = R(t-t') \) and the FDT takes the form
\[ R(t, t') = \beta \frac{\partial}{\partial t'} C(t-t') \theta(t-t') \] (4.55)

The FDT can also be expressed in the frequency domain, where it relates the imaginary part of the response to oscillatory perturbations to the spectral density of the autocorrelation function
\[ \tilde{\chi}''(\omega) = \text{Im} \tilde{\chi}(\omega) = \int_{-\infty}^{\infty} dt R(t) \sin(\omega t) \\
= -\beta \int_{0}^{\infty} dt \frac{d}{dt} \langle \dot{\delta A}(0) \delta A(t) \rangle \sin(\omega t) \\
= \beta \int_{0}^{\infty} dt \langle \dot{\delta A}(0) \delta A(t) \rangle \omega \cos(\omega t) = \frac{\omega}{2} S(\omega) \] (4.56)
where we have used the definition of Fourier transform, the fluctuation theorem (4.39), integration by part, the autocorrelation property \( C(t) = C(-t) \) and the Wiener-Khintchine theorem (4.31). Hence,
\[ C(t) = \frac{1}{\beta} \int \frac{d\omega}{\pi} \frac{\tilde{\chi}''(\omega)}{\omega} e^{-i\omega t} = \frac{1}{\beta} \int \frac{d\omega}{\pi} \frac{\tilde{\chi}''(\omega)}{\omega} \cos(\omega t) \] (4.57)
where we used \( \tilde{\chi}''(\omega) = -\tilde{\chi}''(-\omega) \). The results (4.56) and (4.57) apply to correlation and response of two dynamical variables \( A, B \) that have the same symmetry under time reversal, i.e. such that \( \langle \dot{\delta A}(0) \delta B(t) \rangle = \langle \delta A(t) \delta B(0) \rangle \), so that \( C(t) = C(-t) \) at stationarity.

We will show in the next section that the imaginary part of the response is linked to dissipation.
4.6.3 Power absorption

The work per unit time done by an external force $f(t)$ to change the state of the system $x$ by an amount $dx$ is

$$\frac{dW}{dt} = f(t) \frac{dx}{dt} \quad (4.58)$$

The power absorbed by the system equals the average rate at which work is done on the system

$$P(t) = \langle \frac{dW}{dt} \rangle = f(t) \frac{d}{dt} \int dt' R(t-t') f(t') \quad (4.59)$$

If we rewrite the right hand side in terms of Fourier transforms $f(\omega)$ and $\tilde{\chi}(\omega)$ we obtain

$$P(t) = f(t) \frac{d}{dt} \int d\omega \frac{2\pi}{4\pi^2} e^{-i(\omega+\omega')t} f(\omega') \tilde{\chi}(\omega) f(\omega) \omega \quad (4.60)$$

We can now compute the power absorbed and the total energy absorbed for various types of external forces.

Delta function force

Let us assume that at $t = 0$ a delta force is applied $F(\delta(t))$. Then

$$f(\omega) = f \quad (4.61)$$

Substituting into (4.60), we obtain

$$P(t) = -if^2 \int d\omega d\omega' \frac{2\pi}{4\pi^2} e^{-i(\omega+\omega')t} \tilde{\chi}(\omega) \omega \quad (4.62)$$

We can find the total energy absorbed by integrating over all times

$$W_{abs} = \int_{-\infty}^{\infty} dt P(t) = -if^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \omega \chi(\omega) \quad (4.63)$$

Since $\tilde{\chi}'(\omega) = \tilde{\chi}'(-\omega)$ and $\tilde{\chi}''(-\omega) = -\tilde{\chi}''(\omega)$ only the imaginary part contributes, as it should since the total energy absorbed must be a real quantity

$$W_{abs} = f^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \omega \chi''(\omega) \quad (4.64)$$
Oscillating force

Now let us consider a monochromatic oscillating force with pulsation \( \omega_0 \)

\[
f(t) = f \cos(\omega_0 t) = \frac{f e^{-i\omega_0 t} + e^{i\omega_0 t}}{2}
\]

so that

\[
\tilde{f}(\omega) = \pi f [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]
\]

One has

\[
\langle x(t) \rangle = f [\tilde{\chi}'(\omega_0) \cos(\omega_0 t) + \tilde{\chi}''(\omega_0) \sin(\omega_0 t)]
\]

From experience with forced harmonic oscillator, we know that dissipation is governed by that part of the response which is in quadrature of phase with the driving force (i.e. out of phase by \( \pi/2 \) with the force), so the dissipative part of the response is controlled by the imaginary part \( \tilde{\chi}''(\omega_0) \) of \( \tilde{\chi}(\omega_0) \), whereas its in-phase, reversible, part is controlled by the real part \( \tilde{\chi}'(\omega_0) \) of \( \tilde{\chi}(\omega_0) \). Using the Fourier transform we obtain

\[
P(t) = -\frac{if^2}{4}[\omega_0 \tilde{\chi}(\omega_0)(e^{-2i\omega_0 t} + 1) - \omega_0 \tilde{\chi}(-\omega_0)(e^{2i\omega_0 t} + 1)]
\]

The average over a period \( T = \pi/\omega_0 \) gives

\[
W_{\text{abs}} = -\frac{if^2\omega_0^2}{4\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} [\tilde{\chi}(\omega_0)(e^{-2i\omega_0 t} + 1) - \tilde{\chi}(-\omega_0)(e^{2i\omega_0 t} + 1)]
\]

Using the symmetry properties of \( \tilde{\chi}'(\omega) \) and \( \tilde{\chi}''(\omega) \) we get

\[
W_{\text{abs}} = \frac{f^2\omega_0^2}{2} \tilde{\chi}''(\omega_0)
\]

So the average power absorbed depends on the imaginary part of the response function. In principle, the average power absorbed can be measured and therefore \( \tilde{\chi}''(\omega) \) can be measured. Finally, since the fluctuation-dissipation theorem relates \( \tilde{\chi}''(\omega) \) to the spectral density of equilibrium fluctuations, it is possible to probe equilibrium fluctuations by applying a weak external field to the system.

**Exercise** - Consider a forced one-dimensional harmonic oscillator with mass \( m \), natural frequency \( \omega_0 \) and damping constant \( \gamma \)

\[
\ddot{x}(t) + \frac{\gamma}{m} \dot{x} + \omega_0^2 x = \frac{f(t)}{m}
\]

The equilibrium position of the oscillator, in absence of force (i.e. for \( f(t) = 0 \)), is \( x = 0 \). Assume the average displacement from the equilibrium position in presence of the force \( f(t) \) is

\[
\langle x(t) \rangle = \int dt' R(t - t') f(t')
\]

where \( R(t) \) is the response function. Define the dynamical susceptibility \( \chi(\omega) \) as the Fourier transform of the response function

\[
R(t) = \int \frac{d\omega}{2\pi} \chi(\omega)e^{-i\omega t}
\]
(a) Write the explicit expression of the real and imaginary part of $\chi(\omega)$, $\chi'(\omega)$ and $\chi''(\omega)$ respectively.

(b) Find the location of the poles $\omega_1$ and $\omega_2$ of $\chi(\omega)$ in the complex $\omega$ plane.

(c) Find the response function $R(t)$ in the cases $\gamma/m < 2\omega_0$ (underdamping) and $\gamma/m > 2\omega_0$ (overdamping).

(d) Starting from the work per unit time done by the external force on the oscillator

$$\frac{dW}{dt} = f(t)\dot{x}(t) \quad (4.74)$$

and taking a periodic $f(t)$

$$f(t) = f \cos(\omega_0 t) \quad (4.75)$$

show that the instantaneous power absorbed by the medium is

$$P(t) = \frac{if^2}{4} \left[\omega_0 \chi(\omega_0)(e^{-2i\omega_0 t} + 1) - \omega_0 \chi(-\omega_0)(e^{2i\omega_0 t} + 1)\right] \quad (4.76)$$

(e) Show that the time average over a period $T = \pi/\omega_0$ is

$$W_{\text{abs}} = -f^2 \frac{\omega_0^2 \chi''(\omega_0)}{2} \quad (4.77)$$

(f) By invoking the fluctuation-dissipation theorem find an expression for the correlation function $\langle xx\rangle_{eq}$ for the fluctuations about the equilibrium position in absence of the external force.
Chapter 5

Fokker Planck

5.1 Kramers-Moyal expansion

For the case when \( x \) is a continuous variable and changes in \( x \) take place in small jumps, the
Kramers-Moyal expansion of the master equation casts this integro-differential equation into the
form of a differential equation of infinite order. It is therefore not easier to handle, but under
certain conditions, one may break off after a suitable number of terms. When this is done after
the second-order terms one gets a partial differential equation of second order for \( P_1(x,t) \) called
the Fokker-Planck equation.

Let us first express the transition probability as a function of the size \( \xi = x - x' \) of jumps,
\( W(x|x') = W(x - x'|x') = W(\xi|x') \). The master equation then reads

\[
\frac{\partial}{\partial t} P_1(x,t) = \int d\xi [W(\xi|x-\xi)P_1(x-\xi,t) - W(-\xi|x)P_1(x,t)] - P_1(x,t) \int d\xi W(-\xi|x) \tag{5.1}
\]

where the sign change associated with the change of variable \( x' \rightarrow \xi = x - x' \) is absorbed in the
boundaries.\(^1\)

If we now assume that changes in the coordinate \( x \) can only take place in small jumps, the
transition probability per unit time \( W(x|x') \) decreases rapidly as \( |x-x'| \) increases, i.e. \( W(\xi|x-\xi) \)
is a sharply peaked function of \( \xi \), but varies slowly enough with \( x \). A second assumption is that
\( P_1(x,t) \) itself also varies slowly with \( x \). Now expand \( W(\xi|x-\xi)P_1(x-\xi,t) \) in powers of \( \xi \)

\[
\frac{\partial}{\partial t} P_1(x,t) = \int d\xi [W(\xi|x)P_1(x,t) + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \xi^m \frac{\partial^m}{\partial x^m}[W(\xi|x)P_1(x,t)]] - P_1(x,t) \int d\xi W(-\xi|x) \tag{5.2}
\]

\(^1\)by considering a symmetrical integration interval extending from \(-\infty\) to \(+\infty\)

\[
\int_{-\infty}^{\infty} dx' f(x') = - \int_{x+\infty}^{x-\infty} d\xi f(x-\xi) = - \int_{-\infty}^{\infty} d\xi f(x-\xi) = \int_{-\infty}^{\infty} d\xi f(x-\xi)
\]

Moreover, since finite integration limits would incorporate an additional dependence on \( x \) we shall restrict to
problems to which the boundary is irrelevant.
the first and the last term on the RHS cancel and we are left with
\[
\frac{\partial P_1(x,t)}{\partial t} = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial x^m} \left\{ \left[ \int d\xi \xi^m W(\xi|x) \right] P_1(x,t) \right\}
\]
(5.3)

Finally, on introducing the jump moments
\[
a^{(m)}(x) = \int d\xi \xi^m W(\xi|x)
\]
(5.4)
one gets the Kramers-Moyal expansion of the master equation
\[
\frac{\partial P_1(x,t)}{\partial t} = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial x^m} \left[ a^{(m)}(x,t) P_1(x,t) \right]
\]
(5.5)

Formally, equation (5.5) is identical with the master equation and is therefore not easier to deal with, but it suggests that one may break off after a suitable number of terms. For instance, there could be situations where, for \( m > 2 \), \( a^{(m)}(x,t) \) is identically zero or negligible. In this case one is left with
\[
\frac{\partial P_1(x,t)}{\partial t} = - \frac{\partial}{\partial x} \left[ a^{(1)}(x,t) P_1(x,t) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[ a^{(2)}(x,t) P_1(x,t) \right]
\]
(5.6)
which is the celebrated Fokker-Planck equation. The first term is called drift or transport term and the second one is the diffusion term, while \( a^{(1)}(x,t) \) and \( a^{(2)}(x,t) \) are the drift and diffusion "coefficients". By definition the Fokker-Planck equation is always a linear equation. Hence the adjective “linear” is available for use in a different sense. We shall call a FP equation linear if the drift coefficient is a linear function of \( x \), \( a^{(1)} = A_0 + A_1 x \) and the diffusion coefficient is constant.

It is worth recalling that, as the Master equation, the Fokker-Planck equation applies to \( P^\star_1(x,t) = P_1|x_0,t_0 \) of every subprocess that can be extracted from a Markov stochastic process by imposing initial condition \( P^\star_1(x,t_0) = \delta(x - x_0) \).

It is important to note that in many physical cases the size of the jumps is determined by some discrete finite parameter and the assumption inherent in the derivation of the Fokker-Planck equation does not apply.

### 5.1.1 The jump moments

The main use of FP equation is as an approximate description for any Markov process whose individual jumps are small. Although the FP equation can still not be solved explicitly except for a few special cases, it can be determined for any actual stochastic process with a minimum knowledge about the underlying mechanism: in fact, it does not require the knowledge of the entire kernel \( W(\xi|x) \) but merely of the functions \( a^{(1)} \) and \( a^{(2)} \).

In order to calculate \( a^{(m)}(x,t) \) we must use the relation (2.26) between \( W \) and the transition probability for short time differences. Firstly, we use \( W(x'|x) = W(\xi|x) \) with \( x = x' + \xi \) and rewrite the jump moments as
\[
a^{(m)}(x,t) = \int dx' (x' - x)^m W(x'|x)
\]
(5.7)
We introduce the quantity

\[ M^{(m)}(x; \Delta t, t) = \int dx' (x' - x)^m P(x', t + \Delta t|x, t) \quad (m \geq 1) \]  

(5.8)

which is the average of \([X(t + \Delta t) - X(t)]^m\) with sharp initial value \(X(t) = x\) (conditional average). Then by using the short-time transition probability (2.26), one can write

\[ M^{(m)}(x; \Delta t, t) = \int dx' (x' - x)^m \{ \delta(x' - x)[1 - a^{(0)}(x, t)\Delta t] + W(x'|x)\Delta t + \mathcal{O}(\Delta t^2) \} \]

\[ = \Delta t \int dx' (x' - x)^m W(x'|x) + \mathcal{O}(\Delta t^2) \]

\[ = a^{(m)}(x, t)\Delta t + \mathcal{O}(\Delta t^2) \]  

(5.9)

Therefore one can calculate the jump moments from the derivatives of the conditional averages as follows

\[ a^{(m)}(x, t) = \frac{\partial}{\partial \Delta t} M^{(m)}(x; \Delta t, t)|_{\Delta t=0} \]  

(5.10)

Finally, on writing

\[ M^{(m)}(x, \Delta t, t) = \langle [X(t + \Delta t) - X(t)]^m \rangle|_{X(t)=x} \]  

(5.11)

one can alternatively express the jump moments as

\[ a^{(m)}(x, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle [X(t + \Delta t) - X(t)]^k \rangle|_{X(t)=x} \]  

(5.12)

Therefore, setting up a Fokker-Plank equation only requires computing the average change \(\langle \Delta x \rangle\) and its mean square \(\langle (\Delta x)^2 \rangle\) to first order in \(\Delta t\). In this way the actual equation of motion can be solved only during \(\Delta t\), which can be done by some perturbation theory. The Fokker-Planck equation then serves to find the long-time behaviour. In Chapter 6, which is devoted to Langevin, we shall calculate the corresponding jump moments in terms of the short-time conditional averages by means of this formula.

### 5.1.2 Expression for the multivariate case

The above formulae can be extended to the case of a multi-component Markov process \(X_i(t), i = 1, 2, \ldots, N\). The Kramers-Moyal expansion is in this case given by the multivariate Taylor expansion

\[ \frac{\partial P(x, t)}{\partial t} = \sum_{m=1} \frac{(-1)^m}{m!} \sum_{j_1 \ldots j_m} \frac{\partial^m}{\partial x_{j_1} \ldots \partial x_{j_m}} \left[ a_{j_1 \ldots j_m}^{(m)}(x, t)P(x, t) \right] \]  

(5.13)

while the Fokker-Planck equation is then given by

\[ \frac{\partial P(x, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i} [a_i^{(1)}(x, t)P(x, t)] + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}^{(2)}(x, t)P(x, t)] \]  

(5.14)

The jump moments are given by the natural generalization of (5.4)

\[ a_{j_1 \ldots j_m}^{(m)}(x) = \int dx' (x'_{j_1} - x_{j_1}) \ldots (x'_{j_m} - x_{j_m}) W(x'|x) \]  

(5.15)

and can be calculated by mean of the corresponding generalization of (5.12)

\[ a_{j_1 \ldots j_m}^{(m)}(x, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle \prod_{\mu=1}^m [X_{j_\mu}(t + \Delta t) - X_{j_\mu}(t)] \rangle_{X(t)=x} \]  

(5.16)
Neglecting the fluctuations means, for the Fokker-Planck equation, neglecting the diffusion term. In this case the equation reduces to the Liouville equation

\[
\frac{\partial P(x,t|y,t')}{\partial t} = \sum_i \frac{\partial}{\partial x_i} [a_i^{(1)}(x,t)P(x,t|y,t')] \tag{5.17}
\]

which occurs in classical mechanics. This equation describes a completely deterministic motion, i.e. it is equivalent to the system of differential equations

\[
\frac{dx(t)}{dt} = a^{(1)}(x(t),t) \tag{5.18}
\]

In fact, if a system is governed by (5.18), it follows from Section 4.2 that the equation of motion for the probability \(P(x,t)\) to find the system in state \(x\) is

\[
\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x} \left( \frac{1}{2} \frac{\partial}{\partial x} a^{(2)}(x,t) \right) P(x,t) \tag{5.19}
\]

It is easy to check that if \(z(y,t)\) is the solution of (5.18) with initial condition \(z(y,t') = y\), then the solution to (5.17) with the initial condition

\[
p(x,t'|y,t') = \delta(x-y) \tag{5.20}
\]

is

\[
p(x,t|y,t') = \delta[x - z(y,t)] \tag{5.21}
\]

Thus, if the particle is in a well-defined initial position \(y\) at time \(t'\), it stays on the trajectory obtained by solving the ordinary differential equation.

**Exercise** Prove this result by direct substitution.

**Exercise** Consider a particle moving under the effect of the Hamiltonian \(H(x,p) = p^2/2m + V(x)\). Write the equation of motion for the probability density to find the system in a region of phase space \((x,p)\).

### 5.2 Fokker-Planck as a continuity equation

Equation (5.6) may be written as a continuity equation for the probability density

\[
\frac{\partial P}{\partial t} = -\frac{\partial J}{\partial x} \tag{5.22}
\]

where \(J(x,t)\) is a probability flux or current

\[
J(x,t) = \left[ a^{(1)}(x,t) - \frac{1}{2} \frac{\partial}{\partial x} a^{(2)}(x,t) \right] P(x,t) \tag{5.23}
\]

If the probability current vanishes at the boundaries \(x = x_{\text{min}}\) and \(x = x_{\text{max}}\), (5.22) guarantees that the normalization is preserved at all times

\[
\int_{x_{\text{min}}}^{x_{\text{max}}} dx \ P(x,t) = \text{const} \tag{5.24}
\]
5.3. MACROSCOPIC EQUATION

For natural boundary conditions $x_{\text{min}} = -\infty$ and $x_{\text{max}} = +\infty$, $P(x,t)$ and the probability current (5.23) also vanish at $x = \pm \infty$.

For a stationary process the probability current must be constant. So if the probability current vanishes somewhere, it must be zero everywhere. So with natural boundary conditions the probability current must be zero. So we get the stationary solution from

$$J = a^{(1)} P_s(x) - \frac{1}{2} \frac{\partial}{\partial x} a^{(2)} P_s(x) = 0$$  \hspace{1cm} (5.25)

which we can rewrite as

$$- \frac{\partial}{\partial x} [a^{(2)} P_s] + 2 a^{(1)} \frac{a^{(1)}}{a^{(2)}} [a^{(2)} P_s] = 0$$  \hspace{1cm} (5.26)

Upon integration we obtain

$$P_s(x) = \frac{N_0}{a^{(2)}} e^{2 \int dx' \frac{a^{(1)}}{a^{(2)}}} = N_0 e^{-\phi(x)}$$  \hspace{1cm} (5.27)

where we have defined the potential $\phi(x)$ as

$$\phi(x) = \log a^{(2)} - 2 \int^x dx' \frac{a^{(1)}}{a^{(2)}}$$  \hspace{1cm} (5.28)

and $N_0$ is an integration constant which ensures normalization of $P_s(x,t)$, i.e.

$$P_s(x) = \frac{e^{-\phi(x)}}{\int dx e^{-\phi(x)}}$$  \hspace{1cm} (5.29)

Note that it is not always possible to find an integrable solution. For linear Fokker-Planck equation ($a^{(1)} = A_0 + A_1 x$) with $A_1 < 0$, the stationary solution (5.27) is Gaussian. In fact, the stationary Markov process determined by the linear Fokker-Planck equation is the Ornstein-Uhlenbeck process.

Nonstationary solutions of the FP equation are more difficult to obtain. A general expression for the nonstationary solution can be found only for special drift and diffusion coefficients.

5.3 Macroscopic equation

If we multiply (5.6) by $x$ and integrate over $x$ we find

$$\partial_t \langle x \rangle = \langle a^{(1)}(x) \rangle$$  \hspace{1cm} (5.30)

where we carried out integrations by part and assumed that $P$ and its derivative goes to zero as $x \to \infty$. Similarly, one can get an equation for the fluctuations

$$\frac{\partial}{\partial t} \langle x^2 \rangle = 2 \langle x a^{(1)}(x,t) \rangle + \langle a^{(2)}(x,t) \rangle$$  \hspace{1cm} (5.31)

One can use the macroscopic equation (5.30) as an alternative, more phenomenological way to get the drift and diffusion coefficient, in the case of linear Fokker-Planck equation.
If the equation is linear one has $\langle a^{(1)}(x) \rangle = a^{(1)}(\langle x \rangle)$ and one obtains a differential equation for $\langle x \rangle$

$$\partial_t \langle x \rangle = a^{(1)}(\langle x \rangle)$$

(5.32)

This equation is identified with the macroscopic equation of motion of the system which is probably known. So one can get $a^{(1)}$ from the knowledge of the macroscopic behaviour and subsequently can find $a^{(2)}$ by identifying (5.27) with the equilibrium distribution one can get from statistical mechanics.

### 5.4 Examples of Fokker-Planck equations

#### Wiener process

In Einstein’s explanation of Brownian motion, he arrived at the following diffusion equation for the position

$$\frac{\partial}{\partial t} P = D \frac{\partial^2 P}{\partial x^2}$$

(5.33)

Comparing with (5.6), this is a Fokker-Planck equation with vanishing drift coefficient $a^{(1)} = 0$ (because no forces act on the particle, hence the net drift is zero) and constant diffusion coefficient $a^{(2)}(x) = 2D$ (because the properties of the surrounding medium are homogeneous, otherwise $D = D(x)$).

The solution of this equation for initial condition $P(x,0) = \delta(x)$ was (4), which corresponds to the Wiener-Levy process (2.49).

#### Klein-Kramers equations

The true diffusion equation, in phase space $(x,v)$, for a Brownian particle of mass $m$ moving in a potential $V(x)$, with arbitrary damping coefficient $\gamma$, is

$$\frac{\partial P}{\partial t} = -v \frac{\partial P}{\partial x} + \frac{\partial}{\partial v} \left[ \left( \frac{\gamma}{m} v - \frac{F(x)}{m} \right) P \right] + \frac{\gamma k_B T}{m^2} \frac{\partial^2 P}{\partial v^2}$$

(5.34)

where $F(x) = -dV/dx$. We shall give a proof of (5.34) in the context of Langevin equations (Section 6.7).

#### Ornstein-Uhlenbeck process

We have stated without proof that the Ornstein-Uhlenbeck process describes the time evolution of the transition probability of the velocity of a free Brownian particle. We shall demonstrate this, by solving the equation for the marginal distribution for $v$ obtained from the no potential limit of (5.34). Integrating (5.34) and using $\int dx \partial_x P(x,v,t) = 0$ (since $P(x = \pm \infty, v,t) = 0$) we obtain the equation for the marginal probability $P_V(v,t) = \int dx P(x,v,t)$

$$\frac{\partial P}{\partial t} = \frac{\gamma}{m} \left( \frac{\partial}{\partial v}(vP) + \frac{k_B T}{m^2} \frac{\partial^2 P}{\partial v^2} \right)$$

(5.35)

where we have rescaled $\gamma$ to absorb the factor $m$, i.e. $\gamma \rightarrow \gamma m$. This is a FP equation with linear drift coefficient and constant diffusion coefficient and it is solved by the transition probability of the O-U process (2.53).
5.4. EXAMPLES OF FOKKER-PLANCK EQUATIONS

To demonstrate the usefulness of the FP equation we calculate the stationary distribution function for the Brownian motion process described by equation (5.35). We immediately get

$$J = \left(-\gamma v - \frac{\gamma kT}{m} \frac{\partial}{\partial v}\right)P_s = 0$$  \hspace{1cm} (5.36)

which leads to the Maxwell distribution

$$P_s(v) = \sqrt{\frac{m}{2\pi kT}} \exp\left(-\frac{mv^2}{2kT}\right)$$  \hspace{1cm} (5.37)

Next we turn to nonstationary solution of (5.35) subject to the initial condition

$$P(v, t|v_0, t_0) = \delta(v - v_0)$$  \hspace{1cm} (5.38)

The solution is best found by making a Fourier transform with respect to $v$, i.e.

$$P(v, t|v', t') = \frac{1}{2\pi} \int dk \; \tilde{P}(k, t|v', t')e^{-ikv}$$  \hspace{1cm} (5.39)

If we multiply the Fokker-Planck equation by $e^{ikv}$ and integrate over $v$, by carrying out a few integrations by part and setting to zero the contribution from the boundaries we get an equation for the Fourier transform (effectively one replaces $\partial/\partial v$ with $-ik$ and $v$ with $-i\partial/\partial k$)

$$\frac{\partial \tilde{P}}{\partial t} = -\frac{\gamma}{m} k \frac{\partial}{\partial k} \tilde{P} - \frac{\gamma k_B T}{m^2} k^2 \tilde{P}$$  \hspace{1cm} (5.40)

which is simpler than (5.35) because only first order derivatives with respect to $k$ occur. Because of (5.38), the initial condition for the Fourier transform is

$$\tilde{P}(k, t'|v_0, t_0) = e^{ikv_0}$$  \hspace{1cm} (5.41)

We set $\tau = m/\gamma$ and $D = \kappa_B T/m$, rewrite the first order equation (5.40) as

$$\tau \frac{\partial \tilde{P}}{\partial t} + k \frac{\partial \tilde{P}}{\partial k} = -Dk^2 \tilde{P}$$  \hspace{1cm} (5.42)

and solve it by using the method of characteristics. The subsidiary system is

$$\frac{dt}{\tau} = \frac{dk}{k} = -\frac{d\tilde{P}}{Dk^2\tilde{P}}$$  \hspace{1cm} (5.45)

\footnote{Briefly, if we have a differential equation of the form

$$P \frac{\partial f}{\partial x} + Q \frac{\partial f}{\partial y} = R$$  \hspace{1cm} (5.43)

and $u(x, y, f) = a$ and $v(x, y, t) = b$ are two solution of the subsidiary system

$$\frac{dx}{a} = \frac{dy}{b} = \frac{df}{R}$$  \hspace{1cm} (5.44)

the general solution of the original equation is an arbitrary function of $u$ and $v$, $h(u, v) = 0$.}
Two integrals are easily obtained considering the systems $t, k$ and $k, \tilde{P}$:

\[
\frac{dt}{\tau} = \frac{dk}{k} \rightarrow k = ae^{\frac{t-t_0}{\tau}} \rightarrow ke^{-\frac{t-t_0}{\tau}} = a
\]  \hspace{1cm} (5.46)

\[-Dkd\tilde{k} = \frac{d\tilde{P}}{\tilde{P}} \rightarrow -\frac{1}{2} Dk^2 = \ln \tilde{P} + c \rightarrow e^{Dk^2/2} \tilde{P} = b \]  \hspace{1cm} (5.47)

Then, the solution $h(u, v) = 0$ can be solved for $v$ as $v = \phi(u)$ with still an arbitrary function $\phi$,

\[
\tilde{P}(k, t|v_0, t_0)e^{\frac{Dk^2}{2}} = \phi(ke^{-\frac{t-t_0}{\tau}})
\]  \hspace{1cm} (5.48)

which is to be determined from initial conditions

\[
\tilde{P}(k, t_0|v_0, t_0)e^{\frac{Dk^2}{2}} e^{ikv_0} = \phi(k)
\]  \hspace{1cm} (5.49)

leading to the desired general solution of (5.40)

\[
\tilde{P}(k, t|v_0, t_0) = e^{-\frac{1}{2} Dk^2} \phi(ke^{-(t-t_0)/\tau})
\]  \hspace{1cm} (5.50)

\[
\phi(ke^{-\frac{t-t_0}{\tau}}) = \exp \left[ ikv_0 e^{-(t-t_0)/\tau} - \frac{Dk^2}{2} \left(1 - e^{-2(t-t_0)/\tau}\right) \right]
\]  \hspace{1cm} (5.51)

which is the characteristic function of a Gaussian distribution with $\mu = v_0 e^{-(t-t_0)/\tau}$ and $\sigma^2 = D(1 - e^{-2(t-t_0)/\tau})$. By performing the integral in (5.39) we finally obtain the probability distribution solving (5.42)

\[
P(v, t|v_0, t_0) = \sqrt{\frac{1}{2\pi D(1 - e^{-2(t-t_0)/\tau})}} \exp \left[ \frac{-(v - v_0 e^{-(t-t_0)/\tau})^2}{2D(1 - e^{-2(t-t_0)/\tau})} \right]
\]  \hspace{1cm} (5.52)

and by inserting the original parameters for the original equation for $P_V$

\[
P(v, t|v_0, t_0) = \sqrt{\frac{m}{2\pi \kappa B T(1 - e^{-2\gamma(t-t_0)})}} \exp \left[ \frac{-m(v - v_0 e^{-\gamma(t-t_0)})^2}{2\kappa B T(1 - e^{-2\gamma(t-t_0)})} \right]
\]  \hspace{1cm} (5.53)

In the limit $\gamma \rightarrow 0$ we recover the result for the Wiener process. For positive $\gamma$ and large time differences $\gamma(t-t_0) \gg 1$ the solution crosses over to the stationary Maxwell distribution

\[
P_s(v) = \sqrt{\frac{m}{2\pi k_B T}} e^{-mv^2/2k_BT}
\]  \hspace{1cm} (5.54)

At any arbitrary time it is a Gaussian distribution with mean value $\langle v(t) \rangle = v_0 e^{-\gamma t}$. For $\gamma \leq 0$ no stationary solution exists.

The O-U process may be equally well described by a linear Langevin equation with Gaussian Langevin forces.
Chapter 6

Langevin equation

In this chapter we investigate the solution of the Langevin equation for simple processes. We solve the Langevin equation for Brownian motion and show how quantities of physical interest can be calculated. We also show how a Fokker-Planck equation can be set up starting from the Langevin equation. For the non-linear Langevin equation some difficulties arise when writing the Fokker-Planck equation, leading to the Ito-Stratonovich dilemma.

6.1 Langevin equation for one variable

The Langevin equation for one variable is a differential equation of the form

\[ \frac{dy}{dt} = A(y, t) + B(y, t)\Gamma(t) \]  \hspace{1cm} (6.1)

where \( \Gamma(t) \) is a given stochastic process. The terms \( A(y, t) \) and \( B(y, t)\Gamma(t) \) are often referred to as the drift and diffusion terms, respectively. Due to the presence of \( \Gamma(t) \), equation (6.1) is a stochastic differential equation, i.e. a differential equation comprising terms which are random functions of time, with given stochastic properties. To solve a Langevin equation then means to determine the statistical properties of the process \( y(t) \), which is defined when an initial condition is supplied.

The choice of \( \Gamma \) which renders \( y(t) \) a Markov process is that of white noise\(^1\)

\[
\langle \Gamma(t) \rangle = 0 \\
\langle \Gamma(t)\Gamma(t') \rangle = 2D\delta(t-t') \]  \hspace{1cm} (6.2)

Since (6.1) is a first-order differential equation, for each sample function (realization) of the noise \( \Gamma(t) \), it determines \( y(t) \) uniquely when an initial condition \( y(t_0) \) is given. In addition, the values of the fluctuating term at different times are statistically independent, due to the \( \delta \)-correlated nature of the Langevin force \( \Gamma(t) \). Therefore, the values of \( \Gamma(t) \) at previous times \( t < t_0 \) cannot influence the conditional probability at later times \( t > t_0 \). From these arguments it follows the Markovian character of the solution of the Langevin equation.

\(^1\)We say that \( \Gamma(t) \) is a white-noise force because its spectral density, which by the Wiener-Khintchine theorem is the Fourier transform of the correlation function, is independent of the frequency \( \omega \).
CHAPTER 6. LANGEVIN EQUATION

Since $\Gamma(t)$ is usually the result of a large number of independent processes (e.g. collisions of a Brownian particles with molecules of fluid), it is customary to assume that in addition to (6.2), $\Gamma(t)$ is Gaussian. In other words, one assumes that the random variables $\Gamma_i = \Gamma(t_i)$ are distributed according to the Gaussian distribution function

$$P(\Gamma_1, ..., \Gamma_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det C}} \exp \left[ -\frac{1}{2} \sum_{ij} \Gamma_i (C^{-1})_{ij} \Gamma_j \right]$$  \hspace{1cm} (6.3)$$

with zero mean (i.e. $\langle \Gamma_i \rangle = 0$) and $C_{ij} = \langle \Gamma_i \Gamma_j \rangle$. It then follows that the higher-order correlation functions of $\Gamma(t)$ are obtained from the second order ones by (see Sec. 1.6)

$$\langle \Gamma(t_1) \Gamma(t_2) ... \Gamma(t_{2n-1}) \rangle = 0$$  \hspace{1cm} (6.4)$$

$$\langle \Gamma(t_1) \Gamma(t_2) ... \Gamma(t_{2n}) \rangle = \sum_P \langle \Gamma(t_{i_1}) \Gamma(t_{i_2}) \rangle \langle \Gamma(t_{i_3}) \Gamma(t_{i_4}) \rangle ... \langle \Gamma(t_{i_{2n-1}}) \Gamma(t_{i_{2n}}) \rangle$$  \hspace{1cm} (6.5)$$

where the sum has to be performed over those permutations which lead to different expressions for $C_{i_1 i_2} C_{i_3 i_4} ... C_{i_{2n-1} i_{2n}}$. These are $(2n)! / (2^n n!)$, because there are overall $(2n)!$ permutations of the $2n$ times $t_i$, however the result is invariant under interchange of the two times in each of the $n$ correlation function and under permutation of the $n$ correlation functions.

It should be noted that for singular correlations like (6.2) the distribution function (6.3) makes sense only if the $C_{ij}$ are finite and if the limit $C_{ij} \to 0$ is considered. For instance we may use the representation

$$\delta_\epsilon(t) = \left\{ \begin{array}{ll} \frac{1}{\epsilon} & -\frac{\epsilon}{2} < t < \frac{\epsilon}{2} \\ 0 & \text{elsewhere} \end{array} \right.$$  \hspace{1cm} (6.6)$$

and then consider the limit $\epsilon \to 0$.

6.2 Langevin equation for Brownian motion

Consider the Langevin equation

$$\dot{v} + \gamma v = \Gamma(t)$$  \hspace{1cm} (6.7)$$

with noise correlation function

$$\langle \Gamma(t) \Gamma(t') \rangle = 2D\delta(t - t')$$  \hspace{1cm} (6.8)$$

Because of the linearity of (6.7) it is sufficient to know the two-time correlation (6.8) of the Langevin force to calculate the two-time correlation function of the velocity $\langle v(t_1) v(t_2) \rangle$. If one is interested in multi-time correlation functions of the velocity (or if the equation is not linear) one needs multi-time correlation functions of the Langevin force.

We now proceed to solve equation (6.7) for the initial condition describing a sharp value in $t = 0$, say $v(0) = v_0$:

$$v(t) = v_0 e^{-\gamma t} + \int_0^t \Gamma(t') e^{-\gamma (t-t')} dt'$$  \hspace{1cm} (6.9)$$
By using (6.8) we obtain for the correlation function of the velocity
\[
\langle v(t_1)v(t_2) \rangle = v_0^2 e^{-\gamma(t_1+t_2)} + \int_0^{t_1} \int_0^{t_2} e^{-\gamma(t_1+t_2-t'_1-t'_2)} 2D\delta(t'_1-t'_2) dt'_1 dt'_2
\]
\[
= v_0^2 e^{-\gamma(t_1+t_2)} + \int_0^{\min(t_1,t_2)} e^{-\gamma(t_1+t_2-2t')} 2Dt'
\]
\[
= v_0^2 e^{-\gamma(t_1+t_2)} + \frac{D}{\gamma} (e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)})
\] (6.10)

For large times \(t_1, t_2 \gg 1/\gamma\), the velocity correlation function is independent of the initial velocity \(v_0\) and is only a function of the time difference \(t_1 - t_2\)
\[
\langle v(t_1)v(t_2) \rangle = \frac{D}{\gamma} e^{-\gamma|t_1-t_2|}
\] (6.11)

The unknown constant \(D\) can now be identified by employing the fact that in the stationary state the average energy of the Brownian particle
\[
\langle E \rangle = \frac{1}{2} m \langle v^2(t) \rangle = \frac{mD}{2\gamma}
\]
must have the known thermal value \(\langle E \rangle = kT/2\) given by equipartition law of classical statistical mechanics. Hence
\[
D = \gamma kT/m.
\] (6.12)

This equation relates \(D\), the size of the fluctuating term, to the damping constant \(\gamma\). It is the simplest form of the general fluctuation-dissipation theorem. The noise is fully expressed in the macroscopic damping constant together with the temperature. Just as Einstein relation, one obtains an identity because one knows the size of equilibrium fluctuations from ordinary equilibrium statistical mechanics. The physical picture is that the random kicks tend to spread out \(v\) while the damping term tries to bring \(v\) back to zero. The balance between these two opposing tendencies is the equilibrium distribution.

In practice, the fluctuation-dissipation theorem tells us that in equilibrium, wherever there is damping there must be fluctuations. They are small from the macroscopic point of view because of the factor \(kT\), but can be observed when magnified. For instance, the velocity fluctuations of a Brownian particle are not seen but they build up a mean square displacement which can be observed under a microscope.

### 6.2.1 Mean-square displacement

For the Brownian particle it is difficult to measure the velocity correlation function. It is much easier to study the mean-square value of its displacement.
\[
\langle (x(t) - x_0)^2 \rangle = \langle (\int_0^t v(t_1)dt_1)^2 \rangle = \langle \int_0^t v(t_1)dt_1 \int_0^t v(t_2)dt_2 \rangle = \int_0^t dt_1 \int_0^t dt_2 \langle v(t_1)v(t_2) \rangle
\]
\[
= \left( v_0^2 - \frac{D}{\gamma} \right) \frac{1 - e^{-\gamma t}}{\gamma^2} + \frac{2D}{\gamma^2} t - \frac{2D}{\gamma^3} (1 - e^{-\gamma t})
\] (6.13)

[Note: were we to start with an initial velocity distribution for the stationary state, i.e. \(\langle v_0^2 \rangle = D/\gamma\) rather than with the sharp velocity \(v_0\) the first term on the RHS would vanish.] Note two
interesting limiting cases. For large time, i.e. \( t \gg 1/\gamma \), the leading term is
\[
\langle (x(t) - x_0)^2 \rangle = \frac{2D}{\gamma^2} t = 2D_E t
\]
with \( D_E = \frac{D}{\gamma^2} = \frac{kT}{m\gamma} \) and the particle behaves like a diffusing particle executing a random walk, with Einstein diffusion constant \( D_E \). For \( t \ll 1/\gamma \) one observes a ballistic behaviour
\[
\langle (x(t) - x_0)^2 \rangle = v_0^2 t^2,
\]
i.e. during a short time interval the particle behaves as if it were a free particle moving with the constant thermal velocity \( v = (\kappa T/m)^{1/2} \). Langevin’s derivation is more general than Einstein’s since it also yields the short time dynamics.

### 6.2.2 Stationary velocity distribution function

In order to derive the stationary velocity distribution, we first calculate all the moments \( \langle v_n^2 \rangle \), from which the characteristic function is determined, and then calculate its Fourier transform to obtain the distribution function.

In the stationary state, i.e. for large times, (6.9) becomes (by setting \( t - t' = \tau \) and, because of the factor \( e^{-\gamma \tau} \), extending the range of integration to \( \infty \))
\[
v(t) = \int_0^\infty d\tau e^{-\gamma \tau} \Gamma(t - \tau) \tag{6.14}
\]
Assuming the Langevin force is Gaussian we have
\[
\langle v_{2n+1}(t) \rangle = 0,
\]
\[
\langle v_{2n}(t) \rangle = \int_0^\infty \cdots \int_0^\infty d\tau_1 \cdots d\tau_{2n} e^{-\gamma (\tau_1 + \cdots + \tau_{2n})} \Gamma(t - \tau_1) \cdots \Gamma(t - \tau_{2n})
\]
\[
= \frac{(2n)!}{2^n n!} \left[ \int_0^\infty \int_0^\infty d\tau_1 d\tau_2 e^{-\gamma (\tau_1 + \tau_2)} 2D \delta(\tau_1 - \tau_2) \right]^n = \frac{(2n)!}{2^n n!} \left( \frac{D}{\gamma} \right)^n \tag{6.15}
\]
The characteristic function becomes
\[
G(k) = 1 + \sum_{n=1}^{\infty} (ik)^n \langle v^n(t) \rangle / n! = \sum_{n=0}^{\infty} (ik)^{2n} \langle v^{2n}(t) \rangle / (2n)!
\]
\[
= \exp \left( -\frac{k^2 D}{2\gamma} \right) \tag{6.16}
\]
and the distribution function is therefore given by
\[
P(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(u)e^{-iuv}du = \sqrt{\frac{m}{2\pi kT}} \exp \left( -\frac{mv^2}{2kT} \right) \tag{6.17}
\]
i.e. the stationary distribution is the Maxwell distribution. Indeed, one sees from (6.14) that \( v(t) \) is a linear combination of all values that \( \Gamma(t') \) takes for \( 0 \leq t' \leq t \). Since the joint distribution of all \( \Gamma(t') \) is Gaussian, it follows that the value \( v \) at time \( t \) has a Gaussian distribution. By the same argument, the joint distribution of \( v(t_1), v(t_2), \ldots \) is Gaussian.
6.2.3 Conditional probability

We now wish to calculate the conditional probability $P(v,t|v_0,t_0)$ of finding the particle with velocity $v$ at time $t$, knowing that it had velocity $v_0$ at time $t_0$. We shall take for simplicity $t_0 = 0$ and write $P(v,t|v_0) = P(v,t|v_0)$. First we show that because $\Gamma(t)$ is a Gaussian process, the integral in equation (6.9) is Gaussian. We therefore conclude that the variable $U = v - v_0 e^{-\gamma t}$ is a Gaussian random process. Let us divide the $[0,t]$ interval in $N$ small intervals of length $\epsilon = t/N, N \gg 1$ with $t_n = n\epsilon$ and define the random variable $B^\epsilon_n$ by

$$B^\epsilon_n = \int_{t_n}^{t_n+\epsilon} dt' \Gamma(t')$$  \hspace{1cm} (6.18)

$B^\epsilon_n$ is time independent due to time translation invariance. We note that $\langle B^\epsilon_n \rangle = 0$ and that from (6.8)

$$\langle B^\epsilon_n B^\epsilon_m \rangle = \int dt' dt'' \langle \Gamma(t) \Gamma(t') \rangle = 2D\epsilon \delta_{nm}$$  \hspace{1cm} (6.19)

From (6.9) we have

$$U = v - v_0 e^{-\gamma t} = e^{-\gamma t} \int_0^t dt' \epsilon^{\gamma t'} \Gamma(t')$$

$$= e^{-\gamma t} \sum_{n=0}^{N-1} \int_{t_n}^{t_n+\epsilon} dt' \epsilon^{\gamma t'} \Gamma(t') \simeq e^{-\gamma t} \sum_{n=0}^{N-1} \epsilon^{\gamma t_n} \int_{t_n}^{t_n+\epsilon} dt' \Gamma(t')$$

$$= e^{-\gamma t} \sum_{n=0}^{N-1} \epsilon^{\gamma n} B^\epsilon_n$$  \hspace{1cm} (6.20)

which is a Riemann approximation to the integral in (6.9). This shows that $U$ is the sum of a large number of independent variables. From the central limit theorem the probability distribution of $U$ is Gaussian with mean $\langle U \rangle = 0$ and variance given by $\langle U^2 \rangle$

$$\langle U^2 \rangle = e^{-2\gamma t} \sum_{n,m=0}^{N-1} \epsilon^{\gamma t_n + \gamma t_m} \langle B^\epsilon_n B^\epsilon_m \rangle$$

$$= e2D e^{-2\gamma t} \sum_{n=0}^{N-1} \epsilon^{2\gamma t_n} = 2D e^{-2\gamma t} \int_0^t dt' e^{2\gamma t'}$$

$$= \frac{D}{\gamma} (1 - e^{-2\gamma t}) = \gamma D E (1 - e^{-2\gamma t})$$  \hspace{1cm} (6.21)

This gives the probability distribution of $U$, or equivalently of $v$

$$P(v,t|v_0) = \frac{1}{\sqrt{2\pi\gamma D E (1 - e^{-2\gamma t})}} \exp \left[ -\frac{(v - v_0 e^{-\gamma t})^2}{2\gamma D E (1 - e^{-2\gamma t})} \right]$$  \hspace{1cm} (6.22)

It is instructive to look at the short and long time limits. In the long time limit $t \gg 1/\gamma$, one reaches an equilibrium situation governed by a Boltzmann distribution

$$P(v,t|v_0) \rightarrow \frac{1}{\sqrt{2\pi\gamma D E}} \exp \left( -\frac{v^2}{2\gamma D E} \right) = \sqrt{\frac{m}{2\pi kT}} \exp \left( -\frac{mv^2}{2kT} \right)$$  \hspace{1cm} (6.23)
The limit $t \ll 1/\gamma$

$$P(v,t|v_0) \to \frac{1}{\sqrt{4\pi Dt}} \exp\left(\frac{(v-v_0)^2}{4Dt}\right)$$  \hspace{1cm} (6.24)

shows that the short time limit is dominated by diffusion $\langle (v-v_0)^2 \rangle \sim 2Dt$. As one can write $\langle |v(t+\epsilon) - v(t)| \rangle \propto \sqrt{\epsilon}$, one sees that the trajectory $v(t)$ is a continuous but non-differentiable function of $t$. The stochastic process whose transition probability we have just derived is often called Ornstein-Uhlenbeck process, or just O-U process.

The conditional distribution (6.22) determines uniquely all the higher order distribution functions because the process is Markov.

### 6.3 Colored noise

The process described by the Langevin equation (6.7) with $\delta$-correlated Langevin force is a Markov process, i.e. its conditional probability at time $t_n$ only depends on the value of the stochastic variable $y(t_{n-1})$ at the next earlier time. The Markovian property is destroyed if $\Gamma(t)$ is no longer $\delta$-correlated. For example, the process described by

$$\dot{y} = h(y) + \tilde{\Gamma}(t)$$  \hspace{1cm} (6.25)

with a Gaussian distributed force $\tilde{\Gamma}$ which has correlation

$$\langle \tilde{\Gamma}(t_1)\tilde{\Gamma}(t_2) \rangle = \frac{D}{\gamma} e^{-\gamma|t_1-t_2|}$$  \hspace{1cm} (6.26)

is no longer a Markov process for finite $\gamma$ (although $\tilde{\Gamma}$ itself is still a Markov process). However, if we introduce an additional variable $\eta$, then a Markov process results for the two dimensional process

$$\begin{align*}
\dot{y} &= h(y) + \eta \\
\dot{\eta} &= -\gamma \eta + \Gamma(t)
\end{align*}$$  \hspace{1cm} (6.27)

with $\delta$-correlated noise $\langle \Gamma(t_1)\Gamma(t_2) \rangle = 2D\delta(t_1-t_2)$. Because of (6.7, 6.8, 6.11) it is easily seen that these are equivalent to (6.25, 6.26). Thus, by introducing new random variables, non-Markovian processes may be reduced to Markovian processes.

#### 6.3.1 Derivation of the Fokker-Planck equation

Consider the Brownian motion in the absence of external forces. We are now interested in the time dependence of the probability $P(v,t)$ that the particle’s velocity at time $t$ lies between $v$ and $v+dv$. One expect that this probability does not depend on the entire past history of the particle, but that it is determined if one knows that $v = v_0$ at some earlier time $t_0$. (We make a Markov assumption). Hence, one can write $P$ more explicitly as a conditional probability which depends on $v_0$ and $t_0$ as parameters, $P(v,t|v_0, t_0)$. Since nothing in this problem depends on the origin from which time is measured (i.e. the problem is time translational invariant), $P$ must only depend on the time difference $\tau = t - t_0$. Thus one can simply write $P(v,\tau,v_0)$. If $\tau \to 0$ one knows that $v = v_0$ so

$$P(v,\tau|v) \to \delta(v - v_0)$$  \hspace{1cm} (6.28)
6.3. COLORED NOISE

On the other hand, if $\tau \to \infty$ the particle must come to equilibrium with the surrounding medium at temperature $T$, irrespective of its past history. Hence $P$ becomes independent of $v_0$ and of the time and it must reduce to the canonical Maxwell-Boltzmann distribution

$$ P(v, t|v_0) \to \left(\frac{m\beta}{2\pi}\right)^{1/2} e^{-\frac{\beta m v^2}{2}} $$

(6.29)

We can calculate the moments $D^{(1)}$ and $D^{(2)}$

$$ D^{(1)} = \lim_{\tau \to 0} \frac{1}{\tau} \langle (\Delta v)_v(t) = v \rangle $$

$$ D^{(2)} = \lim_{\tau \to 0} \frac{1}{\tau} \langle (\Delta v)^2_v(t) = v \rangle $$

(6.30)

from the Langevin equation for the Brownian particle (cfr Section 6.2)

$$ \frac{dv}{dt} = -\gamma v + \Gamma(t) $$

(6.31)

$$ \langle \Gamma(t) \Gamma(t') \rangle = 2\gamma kT \frac{m}{\tau} \delta(t - t') $$

(6.32)

By integrating the equation of motion over a small time $\tau$ we get

$$ v(t + \tau) - v(t) = -\gamma \int_t^{t+\tau} v(t') + \int_t^{t+\tau} dt' \Gamma(t') $$

$$ \simeq -\gamma v(t) \tau + \int_t^{t+\tau} dt' \Gamma(t') $$

(6.33)

Taking the ensemble average at fixed $v(t) = v$

$$ \langle \Delta v \rangle = \langle (v(t + \tau) - v(t))_{v(t) = v} \rangle = -\gamma v $$

(6.34)

which leads to

$$ D^{(1)} = -\gamma v $$

(6.35)

Squaring (6.33) and taking the average we also get

$$ \langle (\Delta v)^2 \rangle_{v(t) = v} = \gamma^2 v^2 \tau^2 + \int_t^{t+\tau} dt' dt'' \langle \Gamma(t') \Gamma(t'') \rangle $$

$$ = \gamma^2 v^2 \tau^2 + 2\frac{kT}{m} \frac{\gamma}{\tau} $$

(6.36)

where we have used $\langle \Gamma(t) \rangle = 0$. This yields

$$ D^{(2)} = \frac{2kT}{m} \gamma $$

(6.37)

The Fokker-Planck equation for the Brownian particle is

$$ \frac{\partial P}{\partial \tau} = -\gamma \frac{\partial}{\partial v} (vP) + \frac{kT}{m} \frac{\partial^2 P}{\partial v^2} $$

(6.38)
CHAPTER 6. LANGEVIN EQUATION

This is the FP equation for the O-U process that we have treated in Sec. 5.4.

**Exercise** Show that the property of \( \Gamma(t) \) to be Gaussian white noise is expressed by the following identity of its characteristic functional

\[
\langle \exp[i \int dt \, k(t) \Gamma(t)] \rangle = \exp \left[ -D \int dt \, k^2(t) \right]
\]  

(6.39)

**Exercise** Prove that

\[
W(t) = \int_0^t dt' \, \Gamma(t')
\]

(6.40)

is the Wiener process.

**Answer:**

Let us formally write \( dW/dt = \Gamma(t) \) and see which are the properties of the \( W(t) \) so defined. On integrating over an interval \( \tau \) we get

\[
w(\tau) = \Delta W(t) = W(t + \tau) - W(t) = \int_t^{t+\tau} ds \, \Gamma(s)
\]

(6.41)

Let us show that \( w(\tau) \) is indeed the Wiener process. Firstly, \( w(\tau) \) is Gaussian because \( \Gamma(t) \) is so. Furthermore, it is easy to show that

\[
w(0) = 0 \quad \langle w(\tau) \rangle = 0 \quad \text{and} \quad \langle w(\tau_1)w(\tau_2) \rangle = 2D \min(t_1, t_2)
\]

(6.42)

### 6.4 Quasilinear Langevin equation

Suppose one has a physical system with a nonlinear equation of motion \( \dot{y} = h(y) \) and following the Langevin approach one adds a Langevin term to describe the fluctuations in the system

\[
\dot{y} = A(y) + \Gamma(t), \quad \langle \Gamma(t)\Gamma(t') \rangle = 2D \delta(t - t')
\]

(6.43)

We call this equation “quasilinear” to express the fact that the coefficient of \( \Gamma(t) \) is still a constant. Although its solution cannot be given explicitly it can still be argued that it is equivalent to the quasilinear Fokker-Planck equation

\[
\frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y} A(y)P + D \frac{\partial^2 P}{\partial y^2}
\]

(6.44)

First, it is clear that for each sample function \( \Gamma \) equation (6.44) uniquely determines \( y(t) \) when \( y(0) \) is given. Since the values of \( \Gamma \) at different times are stochastically independent, it follows that \( y(t) \) is Markovian. Hence, it obeys a Master equation which may be rewritten in the Kramer-Moyal form. Now we have to calculate the first and second order jump moments (5.4).

An exact consequence of (6.43) is

\[
\Delta y = \int_t^{t+\Delta t} dt' \, A(y(t')) + \int_t^{t+\Delta t} dt' \, \Gamma(t')
\]

(6.45)
Hence the average with fixed $y$ is
\[
(\Delta y)_{y(t)=y} = A(y) \Delta t + O(\Delta t)^2
\]
(6.46)

That yields the first term in (6.44). Next
\[
\langle (\Delta y)^2 \rangle_{y(t)=y} = \left( \int_t^{t+\Delta t} dt' A(y(t')) \right)^2 + 2 \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \langle A(y(t')) \Gamma(t'') \rangle
\]
\[
+ \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \langle \Gamma(t') \Gamma(t'') \rangle
\]
(6.47)

The first term is $O(\Delta t)^2$ and therefore does not contribute to $a_2$. The last line equals $2D \Delta t$ as before and agrees with (6.44). In the second line expand $A(y(t'))$:
\[
2A(y(t)) \Delta t \int_t^{t+\Delta t} dt'' \langle \Gamma(t'') \rangle + 2A'(y(t)) \int_t^{t+\Delta t} dt' \int_t^{t+\Delta t} dt'' \langle y(t') - y(t) \rangle \Gamma(t'') + ...
\]
(6.48)

The first term vanishes and the second one is $O(\Delta t)^2$, because it is a double integral and does not contain delta functions. By similar arguments one sees that the dots are higher order in $\Delta t$ as well as $\langle (\Delta y)^n \rangle$ for $n > 2$. This proves the equivalence of (6.43) and (6.44).

### 6.5 Non-linear Langevin equation

For one stochastic variable $y$, the general Langevin equation has the form
\[
\dot{y} = A(y, t) + B(y, t) \Gamma(t)
\]
(6.49)

The Langevin force is again assumed Gaussian with zero mean and $\delta$ correlation function
\[
\langle \Gamma(t) \rangle = 0, \quad \langle \Gamma(t) \Gamma(t') \rangle = 2D \delta(t - t')
\]
(6.50)

For time independent function $A$ and $B$ and $B \neq 0$ one can always reduce the equation with multiplicative noise
\[
\dot{y} = A(y) + B(y) \Gamma(t)
\]
(6.51)

to an equation with additive noise by a simple transformation of variable
\[
\eta = \int_y^y \frac{dy'}{B(y')} \equiv f(y), \quad y = f^{-1}(\eta)
\]
(6.52)

One can then rewrite (6.51) in terms of $\eta$
\[
\dot{\eta} = A_1(\eta) + \Gamma(t)
\]
(6.53)

with additive noise force given by
\[
A_1(\eta) = A(y)/B(y) = A(f^{-1}(\eta))/B(f^{-1}(\eta))
\]
(6.54)
which is a “quasilinear” equation, because $A_1$ is a non-linear function of $\eta$, but the coefficient of $\Gamma(t)$ is still a constant. This has been shown above to be equivalent to the FP equation for the probability density $P_1$ for the stochastic variable $\eta$

$$\frac{\partial P_1(\eta,t)}{\partial t} = - \frac{\partial}{\partial \eta} A_1(\eta) P_1 + D \frac{\partial^2 P_1}{\partial \eta^2}$$

(6.55)

where the connection between $P_1(\eta,t)$ and $P_1(y,t)$ follows from (1.14) as

$$P_1(\eta) = P_1(y) B(y).$$

(6.56)

Transforming back to the original $y$ we have

$$\frac{\partial P(y,t)}{\partial t} = - \frac{\partial}{\partial y} A(y) P + D \frac{\partial}{\partial y} B(y) \frac{\partial}{\partial y} B(y) P$$

(6.57)

where we used

$$\frac{\partial}{\partial \eta} = \frac{\partial y}{\partial \eta} \frac{\partial}{\partial y} = B(y) \frac{\partial}{\partial y}$$

(6.58)

Thus the Langevin equation (6.1) is equivalent to the Fokker-Planck equation (6.57); or alternatively

$$\frac{\partial P(y,t)}{\partial t} = - \frac{\partial}{\partial y} \{ A(y) + D B(y) B'(y) \} P + D \frac{\partial^2}{\partial y^2} \{ [B(y)]^2 P \}$$

(6.59)

where we used

$$B(y) \frac{\partial}{\partial y} [B(y) P] + B(y) B'(y) P = \frac{\partial}{\partial y} [B^2 P]$$

(6.60)

However, there is a problem concerning (6.49). This equation, as it stands, has no well-defined meaning. To see this, note that $\Gamma(t)$, because it has no correlation time, can be visualized as a sequence of delta peaks arriving at random times $t_i$. According to (6.49) each delta function in $\Gamma(t)$ causes a jump in $y(t)$. Hence, the value of $y$ at the time that the delta function arrives is undetermined and therefore also the value of $B(y)$. The equation does not specify if one should insert in $B(y)$ the value of $y$ before the jump, after the jump, or perhaps the mean value of these two values. These different options lead to different Fokker-Planck equations.

One cannot answer these questions from a purely mathematical point of view, and prescriptions, like the Ito or Stratonovich definitions, are needed to determine $B(y)$ unambiguously.

Stratonovich opted for the mean value. He reads (6.49) as

$$y(t + \Delta t) - y(t) = A(y(t)) \Delta t + B \left( \frac{y(t) + y(t + \Delta t)}{2} \right) \int_t^{t+\Delta t} dt' \Gamma(t')$$

(6.61)

It can be proved that this choice leads to (6.59). This shows that our naive use of the transformation of variables amounted to opting for the Stratonovich interpretation.

Ito opted for the value of $y$ before the arrival of the delta peak. He reads (6.49) as

$$y(t + \Delta t) - y(t) = A(y(t)) \Delta t + B(y(t)) \int_t^{t+\Delta t} dt' \Gamma(t')$$

(6.62)
and proved that this is equivalent to
\[ \frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y} [A(y)P] + D \frac{\partial^2}{\partial y^2} \{[B(y)]^2P\} \] (6.63)

Apparently this interpretation is not compatible with the familiar way of transforming variables: new transformation laws have to be formulated. We note that (6.63) and (6.59) differ in the drift coefficient. The difference between their coefficients is sometimes called "spurious drift". The Stratonovich integral is the natural choice for an interpretation which assumes \( \Gamma \) is a real noise (not a white noise) with finite correlation time, which is then allowed to become infinitesimally small after calculating measurable quantities.

### 6.5.1 Definition of the Stochastic integral

The previous Section calls for the introduction of stochastic integration. Suppose \( f(t) \) is an arbitrary function of time and \( W(t) \) is the Wiener process. We define the stochastic integral
\[
I = \int_{t_0}^{t} f(t')dW(t')
\] (6.64)
as a kind of Riemann-Stieltjes integral. Namely, we divide the interval \([t_0, t]\) into \( n \) subintervals by means of partitioning points
\[ t_0 \leq t_1 \leq \ldots \leq t_{n-1} \leq t \]
and define intermediate points \( \tau_i \) such that
\[ t_{i-1} \leq \tau_i \leq t_i \]
The stochastic integral is defined as the mean square limit\(^2\) of the partial sum
\[
S_n = \sum_{i=1}^{n} f(\tau_i)[W(t_i) - W(t_{i-1})]
\] (6.65)

It is heuristically quite easy to see that in general the integral defined as the limit of \( S_n \) depends on the particular choice of intermediate points \( \tau_i \).

**Exercise** - Show that if we take, for example, \( f(t) = W(t) \) and choose for all \( i \)
\[ \tau_i = \alpha t_i + (1 - \alpha)t_{i-1} \quad (0 < \alpha < 1) \] (6.66)
we get: \( \langle S_n \rangle = \sum_{i=1}^{n} (t_i - t_{i-1})\alpha = (t - t_0)\alpha \).

The choice of intermediate points characterised by \( \alpha = 0 \), i.e. \( \tau_i = t_{i-1} \) defines the Ito stochastic integral of the function \( f(t) \):
\[
I^{(I)} = \int_{t_0}^{t} f(t')dW(t') = \text{ms} - \lim_{n \to \infty} \sum_{i=1}^{n} f(t_i)[W(t_i) - W(t_{i-1})]
\] (6.67)
\(^2\)We say that the mean square limit of \( X_n \) as \( n \to \infty \) is \( X \) and write \( \text{ms-lim}_{n \to \infty} X_n = X \) if \( \lim_{n \to \infty} \langle (X_n - X)^2 \rangle = 0 \).
The following alternative definition of the integral \( I \), is due to Stratonovich
\[
I^{(S)} = \int_{t_0}^{t} f(t') dW(t') = ms - \lim_{n \to \infty} \sum_{i=1}^{n} \frac{f(t_i) + f(t_{i-1})}{2} [W(t_i) - W(t_{i-1})] \tag{6.68}
\]

**Exercise** - Consider, as an example, \( f(t) = W(t) \). Calculate \( I = \int_{t_0}^{t} W(t') dW(t') \) using Ito and Stratonovich rule.

(a) Show that Ito’s rule gives
\[
I^{(I)} = \frac{1}{2} [W(t)^2 - W(t_0)^2 - (t - t_0)]
\]

Note the difference with the ordinary integration in which the term \((t - t_0)\) would be absent. The reason for this is that \( W(t + \tau) - W(t) \) is order \( \sqrt{t} \), so, in contrast with ordinary integration, terms of second order in \( \Delta W \) do not vanish on taking the limit.

(b) Show that the anomalous term \((t - t_0)\) does not occur in the Stratonovich integral
\[
I^{(S)} = \frac{1}{2} [W(t)^2 - W(t_0)^2] \tag{6.69}
\]

### 6.6 The Kramers-Moyal coefficients for the Langevin equation

Here we show a general method to construct a Fokker-Planck equation for Markov processes obeying equation (6.1) with Gaussian delta correlated noise, directly in terms of the coefficients appearing in the equation of motion. Let us calculate the successive coefficients occurring in the Kramers-Moyal expansion
\[
a^{(n)}(x, t) = \lim_{\tau \to 0} \frac{1}{\tau} \langle [y(t + \tau) - x]^n \rangle_{y(t)=x} \tag{6.70}
\]

where \( y(t + \tau) \ (\tau > 0) \) is a solution of (6.49) which at time \( t \) has the sharp value \( y(t) = x \). We first cast the differential equation (6.49) in an integral form:
\[
y(t + \tau) - x = \int_{t}^{t+\tau} dt' A(y(t'), t') + \int_{t}^{t+\tau} dt' B(y(t'), t') \Gamma(t') \tag{6.71}
\]

and assume that \( A \) and \( B \) can be expanded about \( y(t') = x \)
\[
A(y(t'), t') = A(x, t') + A'(x, t')(y(t') - x) + ... \\
B(y(t'), t') = B(x, t') + B'(x, t')(y(t') - x) + ... \tag{6.72}
\]

where the primes denote partial derivative with respect to \( y \), evaluated at the initial point \( x \). Inserting these expansions in (6.71) we have
\[
y(t + \tau) - x = \int_{t}^{t+\tau} dt' [A(x, t) + B(x, t) \Gamma(t')] \\
+ \int_{t}^{t+\tau} dt' [A'(x, t') + B'(x, t') \Gamma(t')][y(t') - x] + ... \tag{6.73}
\]
For \(y(t') - x\) in the integrand we iterate (6.73) thus obtaining
\[
y(t + \tau) - x = \int_t^{t+\tau} dt' [A(x, t) + B(x, t)\Gamma(t')]
+ \int_t^{t+\tau} dt' [A'(x, t') + B'(x, t')\Gamma(t')] \int_t^{t'} dt'' [A(x, t'') + B(x, t'')\Gamma(t'')] + ...
\] (6.74)

If we now take the average of (6.74) for fixed \(y(t) = x\) and use (6.50) we have
\[
\langle y(t + \tau) - x \rangle = \int_t^{t+\tau} dt' A(x, t') + \int_t^{t+\tau} dt' \int_t^{t'} dt'' A(x, t'')A'(x, t')
+ \int_t^{t+\tau} dt' \int_t^{t'} dt'' B(x, t'')B'(x, t') 2D\delta(t' - t'') + ...
\] (6.75)

An important point of definition arises here. It frequently occurs in the study of stochastic differential equations that in integrals involving delta functions the argument of the delta function is equal to either the upper or the lower limit of integrals, that is we find integrals like
\[
I_1 = \int_{t_1}^{t_2} dt f(t)\delta(t - t_1) \quad \text{or} \quad I_2 = \int_{t_1}^{t_2} dt f(t)\delta(t - t_2)
\] (6.76)

Various conventions can be made concerning the values of these integrals. If one counts all the weight of a delta function at the lower limit of an integral and none of the weight at the upper limit, this yields \(I_1 = f(t_1)\) and \(I_2 = 0\), and amounts to choosing the Ito convention, where increments “point to the future” \(\int_{t_0}^{t_1} f(t)\Gamma(t)dt = f(t_0)\int_{t_0}^{t_1} f(t) dW(t) = \lim_{n\to\infty} \sum_{i=1}^{n} f(t_{i-1})[W(t_i) - W(t_{i-1})]\) (note the interpretation of the integral of \(\Gamma(t)\) as the Wiener process \(W(t)\) is used). If the delta function has half its weight counted in both the limit of an integral, this yields \(I_1 = \frac{1}{2}f(t_1)\) and \(I_2 = \frac{1}{2}f(t_2)\) and amounts to choosing Stratonovich convention of stochastic integration \(\int_{t_0}^{t_1} f(x(t), t)dW(t) = \lim_{n\to\infty} \sum_{i=1}^{n} f(\frac{x(t_i) + x(t_{i-1})}{2}, t_{i-1})[W(t_i) - W(t_{i-1})]\). Here we use Stratonovich convention and get
\[
\langle y(t + \tau) - x \rangle = \int_t^{t+\tau} dt' A(x, t') + \int_t^{t+\tau} dt' \int_t^{t'} dt'' A(x, t'')A(x, t') + D \int_t^{t+\tau} dt' B(x, t')B'(x, t') + ...
\] (6.77)

This is equivalent to take for the \(\delta\) function any representation \(\delta_i(t)\) symmetric around the origin, for example (6.6) and finally take \(\epsilon \to 0\). As a result, e.g. \(I_2 = f(t_2)\int_{t_1}^{t_2} d\tau \delta_i(\tau) = f(t_2) \int_{-t_2/2}^{0} d\tau \frac{1}{\tau} = \frac{1}{2}f(t_2)\)

Finally, on taking the limit \(\tau \to 0\) and considering that only terms of order \(\tau\) contribute to \(a^{(1)}(x, t) = \lim_{\tau\to 0} \frac{1}{\tau} \langle y(t + \tau) - y(t) \rangle \big|_{y(t) = x}\), we finally get
\[
a^{(1)}(x, t) = A(x, t) + DB(x, t)\frac{\partial B(x, t)}{\partial x}
\] (6.78)

Note that the terms not written down in (6.74) vanish in the limit \(\tau \to 0\) because integrals not containing the Langevin force are \(O(t^n)\) where \(n\) is the number of integrals, whereas integrals containing the Langevin force either vanish, if the number of integrals is odd, or are order \(O(t^{n/2})\).
CHAPTER 6. LANGEVIN EQUATION

if the number \( n \) of integrals is even. Using these arguments for the higher order coefficients, we conclude that

\[
a^{(2)}(x, t) = \lim_{\tau \to 0} \frac{1}{\tau} \int_{t}^{t+\tau} dt' \int_{t}^{t+\tau} dt'' B(x, t')B(x, t'')2D\delta(t' - t'') = 2DB^2(x, t) \tag{6.79}
\]

whereas all the coefficients \( a^{(n)} \) for \( n \geq 3 \) vanish.

One can see that the term \( a^{(1)} \) contains, in addition to the deterministic drift \( A(x, t) \), a term which is called the spurious or noise-induced drift

\[
a_{\text{spurious}}^{(1)}(x, t) = D\frac{\partial B(x, t)}{\partial x} B(x, t) \tag{6.81}
\]

This arises from the fact that during a change of \( \Gamma(t) \) also \( y(t) \) changes and therefore \( \langle B(y(t), t) \Gamma(t) \rangle \) is no longer zero.

In conclusion, for Markov stochastic processes determined by the Langevin equation (6.49) with \( \delta \)-correlated Gaussian-distributed Langevin forces, all KM coefficients vanish for \( n \geq 3 \). Therefore, the distribution of probability obeys the Fokker-Planck equation

\[
\frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y_i} \left[ A_i(y, t) + DB_i(y, t)\frac{\partial B(y, t)}{\partial y} \right] P + D \frac{\partial^2}{\partial y_i \partial y_j} \left[ B^2(y, t) P \right] \tag{6.82}
\]

### 6.6.1 Multivariate case

The Langevin equation for a multi-component process \( y \) has the form

\[
\frac{dy}{dt} = A_i(y, t) + \sum_k B_{ik}(y, t)\Gamma_k(t) \tag{6.83}
\]

where \( \Gamma_k \) are white noise terms with

\[
\langle \Gamma_k(t) \rangle = 0, \quad \langle \Gamma_k(t)\Gamma_{l}(t') \rangle = 2D\delta_{kl}\delta(t - t') \tag{6.84}
\]

The generalization of (6.78,6.79,6.80) to the multivariate case are

\[
a^{(1)}_i(y, t) = A_i(y, t) + DB_{ik}(y, t)\frac{\partial B_{lk}(y, t)}{\partial x_j} \tag{6.85}
\]

\[
a^{(2)}_{ij}(y, t) = 2DB_{ik}(y, t)B_{jk}(y, t) \tag{6.86}
\]

\[
a^{(m)}_{j_1...j_m}(y, t) = 0, \quad \text{for } m \geq 3 \tag{6.87}
\]

The corresponding Fokker-Planck equation is

\[
\frac{\partial P(y, t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial y_i} \left[ A^{(1)}_i(y, t) + DB_{ik}(y, t)\frac{\partial B_{lk}(y, t)}{\partial y_j} \right] P \right. \\
+ D \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} \left[ \sum_k B_{ik}(y, t)B_{jk}(y, t) \right] P \tag{6.88}
\]

which is entirely determined by the coefficients of the Langevin equation.
6.7 Diffusion in phase space: Klein-Kramers equation

Let us consider the following generalisation of the Langevin equation (6.7), in order to account for the presence of an external potential $V(x)$

$$m \frac{d^2 x}{dt^2} = -\gamma \frac{dx}{dt} - V'(x) + \Gamma(t) \quad (6.89)$$

This is simply Newton equation augmented by the fluctuating force and may describe the motion in one dimension of a colloidal particle in presence of a potential $V(x)$. We can rewrite the second order differential equation as a pair of first order differential equations

$$\dot{x} = v, \quad \dot{v} = -\frac{V'(x)}{m} - \frac{\gamma}{m} v + \frac{1}{m} \Gamma(t) \quad (6.90)$$

The state of the particle is described by its position $x$ and velocity $v$. To make contact with Hamiltonian systems, we may regard the particle as a Hamiltonian dynamical system, in which the momentum equation is augmented by a friction and a noise term.

Then, comparing with the multivariate Langevin equation (6.83) we identify $\Gamma_x(t) = 0$ and $\Gamma_v(t) = \Gamma(t)$, as well as

$$A_x = v \quad B_{xx} = 0 \quad B_{xv} = 0$$
$$A_v = -\frac{1}{m}(\gamma v + V') \quad B_{vx} = 0 \quad B_{vv} = \frac{1}{m} \quad (6.91)$$

Inserting these results in the general Fokker-Planck equation (6.88) one gets

$$\frac{\partial P}{\partial t} = -\frac{\partial P}{\partial x} \dot{x} + \frac{\partial P}{\partial v} \left[ -\frac{1}{m} (\gamma v + V'(x)) P \right] + \frac{D}{m^2} \frac{\partial^2 P}{\partial v^2} \quad (6.92)$$

Gathering the Hamiltonian terms and identifying $D = \gamma k_B T$ we finally find the famous **Klein-Kramers equation**

$$\frac{\partial P}{\partial t} = -v \frac{\partial P}{\partial x} + \frac{V'}{m} \frac{\partial P}{\partial v} + \frac{\gamma}{m} \left( \frac{\partial}{\partial v} v + \frac{k_B T}{m} \frac{\partial^2}{\partial v^2} \right) P \quad (6.93)$$

**Probability flow**

It is instructive to derive Klein-Kramers equation as the equation of motion for the probability density in phase space, along the lines of Section 4.2. The probability to find the Brownian particle in the interval $x \rightarrow x + dx$ and $v \rightarrow v + dv$, at time $t$, is $\rho(x, v, t) dx dv$. As explained in Sec. 4.2 we may view the probability distribution as a fluid whose density at point $(x, v)$ is given by $\rho(x, v, t)$. Then, applying standard fluid dynamics, we find the following laws for the probability flow:

$$\frac{\partial \rho(t)}{\partial t} = -\frac{\partial(\dot{x} \rho(t))}{\partial x} - \frac{\partial(\dot{v} \rho(t))}{\partial v}$$
$$= -\rho(t) \left[ \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{v}}{\partial v} \right] - \dot{x} \frac{\partial \rho(t)}{\partial x} - \dot{v} \frac{\partial \rho}{\partial v} \quad (6.94)$$
where we let \( \rho(t) = \rho(x, v, t) \). The first term on the right hand side, which did not contribute in the case of Hamiltonian systems, now yields a contribution \( -\gamma \rho/m \), due to the presence of dissipation. Also, friction and noise term provide additional contributions to \( \dot{v}\partial\rho/\partial v \), so the equation for the probability flow is now given by

\[
\frac{\partial \rho(t)}{\partial t} = -L_0 \rho(t) - L_1 \rho(t) \tag{6.95}
\]

where the differential operators \( L_0 \) and \( L_1 \) are defined

\[
L_0 = -\frac{\gamma}{m} + v \frac{\partial}{\partial x} - \frac{V'(x)}{m} \frac{\partial}{\partial v} - \frac{\gamma}{m} v \frac{\partial}{\partial v} \tag{6.96}
\]

and

\[
L_1 = \frac{1}{m} \Gamma(t) \frac{\partial}{\partial v} \tag{6.97}
\]

Since \( \Gamma(t) \) is a stochastic variable, the time evolution of \( \rho(x, v, t) \) will be different for each realization of \( \Gamma(t) \). So we are interested in the average probability

\[
P(x, v, t) = \langle \rho(x, v, t) \rangle_{\Gamma} \tag{6.98}
\]

Let us derive \( P(x, v, t) \). This is easier done in terms of the new probability density \( \sigma(t) \), such that

\[
\rho(t) = e^{-L_0 t} \sigma(t) \tag{6.99}
\]

which satisfies the equation

\[
\frac{\partial \sigma}{\partial t} = -V(t) \sigma(t) \tag{6.100}
\]

where

\[
V(t) = e^{L_0 t} L_1(t) e^{-L_0 t} \tag{6.101}
\]

The formal solution is

\[
\sigma(t) = e^{-\int_0^t dt' V(t')} \sigma(0) \tag{6.102}
\]

We now expand the exponential

\[
\sigma(t) = \left[ \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left( \int_0^t dt' V(t') \right)^n \right] \sigma(0) \tag{6.103}
\]

Because the noise \( \Gamma(t) \) has zero mean and is Gaussian, Wick's theorem applies. Only even values of \( n \) remain

\[
\langle \sigma(t) \rangle_{\Gamma} = \left[ \sum_{n=0}^{\infty} \frac{1}{2n!} \left( \int_0^t dt' V(t') \right)^{2n} \right] \sigma(0) \tag{6.104}
\]

and the average \( \langle \left( \int_0^t dt' V(t') \right)^{2n} \rangle_{\Gamma} \) will decompose into \( (2n!) / n! 2^n \) identical terms, each containing a product of \( n \) pairwise averages \( \langle \int_0^t dt_i V(t_i) \int_0^t dt_j V(t_j) \rangle_{\Gamma} \). Thus, we have

\[
\langle \sigma(t) \rangle_{\Gamma} = \left[ \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{1}{2} \int_0^t dt_2 dt_1 \langle V(t_2) V(t_1) \rangle_{\Gamma} \right)^n \right] \sigma(0) \tag{6.105}
\]
6.7. DIFFUSION IN PHASE SPACE: KLEIN-KRAMERS EQUATION

We can now sum this series to obtain

\[ \langle \sigma(t) \rangle_\Gamma = \exp \left[ \frac{1}{2} \int_0^t dt_2 dt_1 \langle V(t_2) V(t_1) \rangle_\Gamma \right] \sigma(0) \]  

(6.106)

Let us compute the integral

\[ \frac{1}{2} \int_0^t dt_2 dt_1 \langle V(t_2) V(t_1) \rangle_\Gamma = \frac{1}{2m^2} \int_0^t dt_2 dt_1 \langle \Gamma(t_2) \Gamma(t_1) \rangle_\Gamma e^{L_0 t_2} \frac{\partial}{\partial v} e^{-L_0 (t_2 - t_1)} \frac{\partial}{\partial v} e^{-L_0 t_1} \]

\[ = \frac{D}{m^2} \int_0^t dt_1 e^{L_0 t_1} \frac{\partial^2}{\partial v^2} e^{-L_0 t_1} \]  

(6.107)

Taking the time derivative of (6.106) and substituting (6.107) we get

\[ \frac{\partial \langle \sigma(t) \rangle}{\partial t} = \frac{D}{m^2} e^{L_0 t} \frac{\partial^2}{\partial v^2} e^{-L_0 t} \langle \sigma(0) \rangle \]  

(6.108)

Going back to the original variables we finally obtain

\[ \frac{\partial \langle \rho(t) \rangle}{\partial t} = -L_0 \langle \rho(t) \rangle_\Gamma + \frac{D}{m^2} \frac{\partial^2 \langle \rho(t) \rangle_\Gamma}{\partial v^2} \]  

(6.109)

The equation for the probability density \( P = \langle \rho \rangle \) takes the form

\[ \frac{\partial P}{\partial t} = -v \frac{\partial P}{\partial x} + \frac{\partial}{\partial v} \left[ \left( \frac{\gamma v}{m} + \frac{V'(x)}{m} \right) P \right] + \frac{D}{m^2} \frac{\partial^2 P}{\partial v^2} \]  

(6.110)

6.7.1 The strong friction limit: Smoluchowski equation

Consider a Brownian particle moving in one dimension in a potential well \( V(x) \), and assume that the friction coefficient \( \gamma \) is very large so that the velocity of the Brownian particle relaxes to its stationary state very rapidly. Then we can neglect time variations in the velocity (set \( dv/dt \approx 0 \)). With this assumption, the Langevin equation reduces to

\[ \dot{x} = -\frac{V'(x)}{\gamma} + \frac{\Gamma(t)}{\gamma} \]  

(6.111)

Comparing with the univariate Langevin equation, we identify

\[ A = -\frac{V'}{\gamma}, \quad B = \frac{1}{\gamma} \]  

(6.112)

Inserting these results in the Fokker-Planck equation one gets (setting \( D/\gamma = \kappa B T \)) the Smoluchowski equation

\[ \frac{\partial P}{\partial t} = \frac{1}{\gamma} \frac{\partial}{\partial x} \left[ \left( V'(x) + \kappa B T \frac{\partial}{\partial x} \right) P \right] \]  

(6.113)

The result \( D/\gamma = \kappa B T \) can be obtained on inserting the marginal Boltzmann distribution \( P_0 \propto \exp[-V(x)/\kappa B T] \) and finding the conditions for it to be a stationary solution. Let us specialize the above results to the case of free particle and harmonic oscillator.
• In absence of potential, Smoluchowski equation reduces to the equation for free diffusion

\[
\frac{\partial P}{\partial t} = \frac{\kappa_B T}{\gamma m} \frac{\partial^2}{\partial x^2} P = D_E \frac{\partial^2}{\partial x^2} P
\]  

(6.114)

with solution given by the Wiener-Levy process.

• For the overdamped harmonic oscillator, \( V(x) = m\omega^2 x^2/2 \), the Fokker-Planck equation is solved by the Ornstein-Uhlenbeck process with \( \tau = \gamma/m\omega^2 \) and \( D = \kappa_B T/m\omega^2 \)

\[
P(x,t) = \sqrt{\frac{m\omega^2}{2\pi\kappa_B T(1-e^{-2t/\tau})}} \exp\left[ -\frac{m\omega^2(x-x_0 e^{-t/\tau})^2}{2\kappa_B T(1-e^{-2t/\tau})} \right]
\]  

(6.115)

6.8 Relaxation times

The solution of the Fokker-Planck equation for the overdamped harmonic oscillator relaxes to the equilibrium state \( P_0 \propto \exp(-x^2/(2\kappa_B T)) \) in the timescale \( \tau = \gamma/m\omega^2 \). This is also called the relaxation time and in this problem it is independent of the temperature.

It is very common in physics to find expressions for different relaxation times that depend exponentially on the temperature \( T \) (Arrhenius law), which is the generic behaviour when, to establish equilibrium at low temperature, potential barriers need to be overcome. This was not the situation in the previous example so such dependence was absent. We shall now solve a simple problem with potential barriers in the regime of low temperature and see how the exponential dependence arises. Let us consider an overdamped particle, described by Smoluchowski equation

\[
\frac{\partial P}{\partial t} = \frac{1}{\gamma} \left( \frac{\partial}{\partial x} V' + \frac{\kappa_B T}{m} \frac{\partial^2}{\partial x^2} \right) P = -\frac{\partial J}{\partial x}
\]  

(6.116)

where the last equality defines the current probability \( J \). At very low temperature, the probability to escape from the metastable minimum is very low (zero at \( T = 0 \), for a deterministic system). Therefore the flux of particles over the barrier is very slow, and we can solve the problem as if it were stationary. Then the expression for \( J \) is assumed to be independent of \( x \), and we can integrate the differential equation for \( P \)

\[
-\gamma J = V'P + \frac{\kappa_B T}{m} \frac{\partial P}{\partial x}
\]  

(6.117)

thus obtaining \((U = mV)\) and \( D = \kappa_B T/m\gamma \)

\[
P = e^{-U/k_B T} \left[ P(c) e^{U(c)/k_B T} - \frac{J}{D} \int_c^x dx' e^{U(x')/k_B T} \right]
\]  

(6.118)

where \( c \) is an arbitrary point. If we choose \( c \) well outside the barrier region \( c \to \infty \), we have \( P(c) \approx 0 \) and we find

\[
J = \frac{DP(x)}{\int_c^x dx' e^{U(x')/k_B T}} e^{U(x)/k_B T}
\]  

(6.119)

Since \( J \) is independent of \( x \), we can choose \( x \) at will. We choose \( x = a \), the metastable minimum, so the integral in the denominator covers the entire maximum. The main contribution to the
6.9. FLUCTUATION-DISSIPATION RELATIONS FOR LANGEVIN EQUATION

The integral comes from a small region about the maximum \( x = b \), so we expand there \( U(x) \simeq U_b - \frac{1}{2} m \omega_b^2 (x - b)^2 \), being \( m \omega_b^2 = U''(b) \). The integration limits can be shifted to \( \pm \infty \), so the resulting Gaussian integral leads to

\[
J = DP(a) \sqrt{\frac{m \omega_b^2}{2 \pi \kappa_B T}} e^{-(U_a - U_b)/\kappa_B T} \tag{6.120}
\]

Since the flux \( J \) leaving the well is given by the number of particles \( N_a \) in the well times the escape rate, the relaxation rate is defined as

\[
\frac{1}{\tau} \equiv \frac{J}{N_a} = D \sqrt{\frac{m \omega_b^2}{2 \pi \kappa_B T}} \frac{P(a)}{N_a} e^{-(U_b - U_a)/\kappa_B T} \tag{6.121}
\]

To obtain the number of particles in the well, we integrate \( P(x) \) in a region around \( a \), with \( P(x) = C e^{-U(x)/\kappa_B T} \) and \( U(x) = U_a + \frac{1}{2} m \omega_a^2 (x - a)^2 \), where \( m \omega_a^2 = U''(a) \), so \( P(a)/N_a = \sqrt{m \omega_a^2 / 2 \pi \kappa_B T} \). So one finds for the relaxation time,

\[
\frac{1}{\tau} = \frac{\omega_a \omega_b}{2 \pi \gamma} e^{-(U_b - U_a)/\kappa_B T} \tag{6.122}
\]

the typical exponential dependence on the barrier height over the temperature.

### 6.9 Fluctuation-dissipation relations for Langevin equation

In this section we derive the fluctuation-dissipation relations in systems subject to a deterministic drift \( \partial H[x(t)]/\partial x(t) \) and a random force \( \Gamma(t) \), evolving according to Langevin dynamics

\[
\frac{dx(t)}{dt} = -\frac{\partial H[x(t)]}{\partial x(t)} + \Gamma(t) \tag{6.123}
\]

where \( \Gamma(t) \) is a Gaussian white noise with \( \langle \Gamma(t) \Gamma(t') \rangle = 2 T \delta(t - t') \). We define the two-time correlation function as

\[
C(t, t') = \langle x(t)x(t') \rangle \tag{6.124}
\]

and assume \( t > t' \). In order to calculate the response of the system to a time-dependent external perturbation, we consider the following perturbation

\[
H_h[x(t)] = H[x(t)] + h(t)x(t) \tag{6.125}
\]

The variation of the quantity \( x(t) \) induced by the presence of \( h(t) \), is measured by the two-time response function \( R(t, t') \), defined as

\[
R(t, t') = \frac{\delta \langle x(t) \rangle}{\delta h(t')} \tag{6.126}
\]

If the system is in equilibrium, the response function is not independent of the correlation, but it is related to the correlation by the celebrated fluctuation-dissipation theorem

\[
R(t, t') = \frac{1}{T} \theta(t - t') \frac{\partial C(t, t')}{\partial t'} \tag{6.127}
\]
and it is also a homogeneous function of time.

Let us compute the following derivative of \( C(t, t') \):

\[
\frac{\partial C(t, t')}{\partial t} = \left\{ \frac{\partial H[x(t)]}{\partial x(t')} + \Gamma(t) \right\} x(t')
\] (6.128)

and subtract the derivative with respect to the argument \( t' \)

\[
\frac{\partial C(t, t')}{\partial t} - \frac{\partial C(t, t')}{\partial t'} = A(t, t') - \langle x(t) \Gamma(t') \rangle \quad \text{with}
\]

\[
A(t, t') = -\langle \frac{\partial H[x(t)]}{\partial x(t)} x(t') \rangle + \langle \frac{\partial H[x(t')]}{\partial x(t')} x(t) \rangle
\] (6.130)

where we dropped the noise term \( \langle x(t') \Gamma(t) \rangle \) because of causality. At equilibrium, due to time reversal, the correlation functions satisfy

\[
\langle B(t) D(t + \tau) \rangle = \langle B(t + \tau) D(t) \rangle
\] (6.131)

where \( B(t), D(t) \) are any two functions of \( x(t) \). Therefore, the asymmetry term \( A \) on the right hand side of (6.129) cancel. The last term can be evaluated by mean of Novikov’s theorem, which states that if \( \Gamma(t) \) is a Gaussian process and \( f \) is a function of \( x(t) \), then

\[
\langle f[x(t)] \Gamma(t') \rangle = \int dt'' \frac{\delta(f[x(t)])}{\delta \Gamma(t'')} \langle \Gamma(t'') \Gamma(t') \rangle
\] (6.132)

Setting \( f[x(t)] = x(t) \), we obtain

\[
\langle x(t) \Gamma(t') \rangle = \int dt'' R(t, t'') \langle \Gamma(t'') \Gamma(t') \rangle = 2TR(t, t')
\] (6.133)

and we find

\[
\frac{\partial C(t, t')}{\partial t} - \frac{\partial C(t, t')}{\partial t'} = -2TR(t, t')
\] (6.134)

At equilibrium, due to \textit{invariance under translation in time}, the two-time average depends only on the time difference,

\[
\frac{\partial C(t, t')}{\partial t} = -\frac{\partial C(t, t')}{\partial t'}
\] (6.135)

hence we retrieve the fluctuation-dissipation theorem

\[
\frac{\partial C(t, t')}{\partial t'} = TR(t, t')
\] (6.136)

and the response function has the form

\[
R(t, t') = \frac{1}{T} \theta(t - t') \frac{\partial C(t, t')}{\partial t'}
\] (6.137)
Proof of Novikov’s theorem

Novikov’s theorem states that for a multivariate Gaussian distribution with zero mean

\[ P(x) = \sqrt{\det A} \exp \left( -\frac{1}{2} x A x \right) \] (6.138)

the averages of the type \( \langle x_i f(x) \rangle \) can be obtained as

\[ \langle f(x) x_i \rangle = \sum_m \left( \delta f / \partial x_m \right) \langle x_i x_m \rangle \] (6.139)

Note that applying this result to \( f(x) = x_j x_k x_\ell \) and using \( \partial x_i / \partial x_m = \delta_{im} \) we get Wick’s formula

\[ \langle x_i x_j x_k x_\ell \rangle = \langle x_i x_j \rangle \langle x_k x_\ell \rangle + \langle x_i x_k \rangle \langle x_j x_\ell \rangle + \langle x_i x_\ell \rangle \langle x_j x_k \rangle \] (6.140)

We prove this theorem in three steps.

- Denote \( E(x) = \frac{1}{2} \sum_{ij} x_i A_{ij} x_j \). We have (using the symmetry of \( A \))

\[ \frac{\partial E}{\partial x_m} = \sum_j A_{mj} x_j \quad \rightarrow \quad x_i = \sum_m (A^{-1})_{im} \frac{\partial E}{\partial x_m} \] (6.141)

- Inserting in the definition of average for \( \langle x_i f(x) \rangle \)

\[ \langle x_i f(x) \rangle = C \int dx x_i f(x) e^{-E(x)} = C \sum_m (A^{-1})_{im} \int dx f(x) \frac{\partial E}{\partial x_m} e^{-E(x)} \]

\[ = C \sum_m (A^{-1})_{im} \int dx \frac{\partial f}{\partial x_m} e^{-E(x)} = \sum_m (A^{-1})_{im} \frac{\partial f}{\partial x_m} \] (6.142)

- Using the above result for \( f(x) = x_j \), and \( \partial x_j / \partial x_m = \delta_{jm} \), we get \( \langle x_i x_j \rangle = (A^{-1})_{ij} \).

Insertion of this above completes the proof.

6.9.1 Fluctuation-dissipation violation

If the system is out of equilibrium, neither homogeneity nor fluctuation-dissipation theorem hold and the asymmetry term may be present. Equation (6.137) is not valid in general and a generalised relation between response and correlation must be considered

\[ R(t, t') = \frac{1}{T} \delta(t - t') X(t, t') \frac{\partial C(t, t')}{\partial t'} \] (6.143)

with \( X \) a function of both times \( t, t' \). Deviations from \( X = 1 \) quantify the out-of-equilibrium dynamics.
The off-dynamics of a simple unfrustrated system

The simplest example of a dynamical system that does not reach equilibrium is the random walk. In the continuum limit the quantity $x(t)$ satisfies the very simple differential equation

$$\frac{dx(t)}{dt} = \Gamma(t)$$

(6.144)

with $\Gamma(t)$ a Gaussian noise with variance given by

$$\langle \Gamma(t)\Gamma(t') \rangle = 2T\delta(t-t')$$

(6.145)

The correlation function is easily obtained from

$$C(t,t') = \langle x(t)x(t') \rangle = \int_0^t dt_1 \int_0^{t'} dt_2 \langle \Gamma(t_1)\Gamma(t_2) \rangle = 2T\min(t,t')$$

(6.146)

whereas in order to get the response function we apply a small perturbation to the Hamiltonian $H \rightarrow H + \int dt h(t)x(t)$ so the dynamical equation becomes

$$\frac{dx(t)}{dt} = h(t) + \Gamma(t)$$

(6.147)

Averaging over the zero mean noise we get

$$\langle x(t) \rangle = \int_0^t dt_1 h(t_1)$$

(6.148)

so

$$R(t,t') = \frac{\delta\langle x(t) \rangle}{\delta h(t')} = \theta(t-t')$$

(6.149)

Hence the relation (6.129) is satisfied with

$$X(t,t') = 1$$

(6.150)

a constant function $\forall t,t'$, but different from the usual FDT result $X = 1$, i.e. the system never reaches equilibrium.
Chapter 7

Applications to complex systems

[This chapter follows closely the book “Neural information processing systems”\(^1\)]

Statistical mechanics deals with large systems of interacting microscopic elements. Its strategy is to avoid solving models of such systems at the microscopic level, but trying instead to use the microscopic laws to calculate laws describing the behaviour of suitable chosen macroscopic observables. It turns out that if we consider large systems, macroscopic laws are usually deterministic and simple even though the microscopic ones are stochastic and complex. Macroscopic observables normally present a kind of average behaviour of the microscopic ones so the more elements we average over the smaller the fluctuations of the average. In the limit of an infinitely large system the fluctuations vanish and the averages evolve deterministically. Statistical mechanics enable the reduction from the microscopic to the macroscopic level. Here we apply statistical mechanics tools to recurrent neural networks (there are feedback loops between the neurons), functioning as associative memories: by appropriate choice of the synaptic weights we can ensure that the neural firing patterns, starting from a given initial state, will converge towards one of a number of patterns that we wish to store.

7.1 Neural networks models

One can think of neural networks as systems of \(N\) neurons which are either firing (on) or quiet (off), so that the state of each neuron can be described by a binary variable \(\sigma_i \in \{-1, 1\}\). Each neuron interacts with other neurons via synaptic connections, whose strengths are represented by real valued numbers \(J_{ik}\), and is influenced by a field at its location

\[
h_i(t) = \sum_{k=1}^{N} J_{ik} \sigma_k(t) + \theta_i
\]

where \(\theta_i\) is a threshold and/or external stimuli, characteristic for each neuron.

In the McCulloch-Pitts model time is discretized in units \(\Delta\) and each neuron evolves in time

\(^{1}\)by ACC Coolen, R Kuhn, P Sollich
according to the deterministic rule

\[ \sigma_i(t + \Delta) = \text{sgn}(h_i(t)) \]  

(7.2)

In the case we wish to take into account noise in systems with McCulloch-Pitts type neurons (neurons not always do what they are expected to do), it is convenient to assume that the thresholds are random variables, distributed symmetrically around an average value \( \theta^*_i \), all with the same variance, and write them in a form where the average and variance are explicit:

\[ \theta_i(t) = \theta^*_i + T z_i(t) \]  

(7.3)

\[ \langle z_i(t) \rangle = 0, \quad \langle z_i^2(t) \rangle = 1 \]  

(7.4)

\[ T^2 = \langle (\theta_i(t) - \theta^*_i)^2 \rangle \]  

(7.5)

The parameter \( T \) measures the level of noise in the system (plays the role of temperature in magnetic systems). For \( T = 0 \) we return to our previous deterministic law; for \( T \to \infty \) the system behaves in a completely random manner. The probability to find a neuron in state \( \sigma_i(t + \Delta) \) can be expressed in terms of the distribution \( P(z) \) of the independent noise variables \( z_i(t) \). For symmetric noise distributions, that is \( P(z) = P(-z) \quad \forall z \), this probability can be written in a very compact way

\[ \text{Prob}[\sigma_i(t + \Delta)] = g\left(\frac{\sigma_i(t + \Delta)h_i(t)}{T}\right) \]  

(7.6)

with

\[ g(x) = \int_{-\infty}^{x} dz P(z) \]  

(7.7)

[We used \( \text{Prob}[\sigma_i(t + \Delta) = 1] = \text{Prob}[h_i(t) + T z_i(t) > 0] = g[h_i(t)/T] \) and \( \text{Prob}[\sigma_i(t + \Delta) = -1] = \text{Prob}[h_i(t) + T z_i(t) < 0] = g[-h_i(t)/T] \)]

A natural choice for the distribution \( P(z) \) of the noise variable \( z_i \) is the Gaussian rule

\[ P(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \]  

(7.9)

which leads to

\[ g(x) = \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \]  

(7.10)

where \( \text{erf}(x) \) denotes the error function. An alternative choice which simplifies considerably many calculations, is to replace the above function \( g(x) \) by the following (qualitatively similar) one:

\[ g(x) = \frac{1}{2} [1 + \tanh(x)] \]  

(7.11)

which corresponds to the noise distribution

\[ P(z) = \frac{1}{2} \left[ 1 - \tanh^2(z) \right] \]  

(7.12)

In a network of such units, updates can be carried out either in parallel (synchronously) or sequentially (one after the other). In the first case, since all noise variables are independent, the
combined probability to find state \( \sigma(t+\Delta) = (\sigma_1(t+\Delta),...,\sigma_N(t+\Delta)) \in \{-1,1\}^N \) equals the product of the individual probabilities, so

\[
\text{parallel : } \mathrm{Prob}[\sigma(t+\Delta)] = \prod_{i=1}^{N} g \left( \frac{\sigma_i(t+\Delta) h_i(t)}{T} \right)
\] (7.13)

In the second case we have to take into account that only one neuron changes its state at a time. The candidate is drawn at random with probability \( N^{-1} \), resulting in

\[
\text{sequential : } \left\{ \begin{array}{l}
\text{choose } i \text{ randomly from } \{1,...,N\} \\
\mathrm{Prob}[\sigma_i(t+\Delta)] = g \left( \frac{\sigma_i(t+\Delta) h_i(t)}{T} \right)
\end{array} \right.
\] (7.14)

If we choose (7.11) for \( g \), abbreviate the inverse noise level as \( T^{-1} = \beta \) and use \( \tanh(\sigma z) = \sigma \tanh(z) \) for \( \sigma \in \{-1,1\} \), we can write the microscopic dynamics in the following form

\[
\text{parallel : } \mathrm{Prob}[\sigma(t+\Delta)] = \prod_{i=1}^{N} \left[ \frac{1}{2} \left[ 1 + \sigma_i(t+\Delta) \tanh(\beta h_i(\sigma(t))) \right] \right]
\] (7.15)

\[
\text{sequential : } \left\{ \begin{array}{l}
\text{choose } i \text{ randomly from } \{1,...,N\} \\
\mathrm{Prob}[\sigma_i(t+\Delta)] = \frac{1}{2} \left[ 1 + \sigma_i(t+\Delta) \tanh(\beta h_i(\sigma(t))) \right]
\end{array} \right.
\] (7.16)

Note that the microscopic laws can be rewritten in a number of ways, for example via

\[
\frac{1}{2} \left[ 1 + \sigma \tanh(\beta h) \right] = \frac{e^{\beta h}}{2 \cosh(\beta h)} = \frac{1}{1 + e^{-2\beta \sigma}}
\] (7.17)

7.2 Microscopic dynamics in probabilistic form

Next, we write the above two versions of the stochastic dynamics in terms of the evolving microscopic state probabilities

\[
p_t(\sigma) = \mathrm{Prob}[\sigma(t) = \sigma]
\] (7.18)

7.2.1 Parallel dynamics

For parallel dynamics we can set the elementary time unit to \( \Delta = 1 \). Then,

\[
p_{t+1}(\sigma) = \prod_{i=1}^{N} \left[ \frac{1}{2} \left[ 1 + \sigma_i(t) \tanh(\beta h_i(\sigma(t))) \right] \right] = \prod_{i=1}^{N} \frac{e^{\beta \sigma_i h_i(\sigma(t))}}{2 \cosh(\beta h_i(\sigma(t)))}
\] (7.19)

If instead of the precise microscopic state \( \sigma(t) \), the probability distribution \( p_t(\sigma') \) over all possible values \( \sigma' \) of \( \sigma(t) \) at time \( t \) is given, the above expression generalizes to the corresponding average over \( \sigma' \):

\[
p_{t+1}(\sigma) = \sum_{\sigma'} W(\sigma,\sigma') p_t(\sigma')
\] (7.20)

This defines a Markov chain with transition probability matrix

\[
W(\sigma,\sigma') = \prod_{i=1}^{N} \frac{e^{\beta \sigma_i h_i(\sigma')}}{2 \cosh(\beta h_i(\sigma'))}
\] (7.21)
which satisfies
\[ W(\sigma, \sigma') \in [0, 1] \quad \sum_{\sigma} W(\sigma, \sigma') = 1 \] (7.22)

If we think of \( p_t(\sigma) \) as a \( 2^N \)-dimensional vector, indexed by the possible states \( \sigma \) of the network, then the vector \( p_{t+1} \) is the product of the \( 2^N \times 2^N \) transition matrix \( W \) with the vector \( p_t \).

### 7.2.2 Sequential dynamics

In the case of sequential dynamics the stochasticity of the dynamics is both in the choice of the site \( i \) to be updated and in the stochastic update of the chosen neuron. If both \( \sigma(t) \) and the chosen site \( i \) are given we find

\[ p_{t+\Delta}(\sigma) = \frac{1}{2} \left[ 1 + \sigma_i \tanh(\beta h_i(\sigma(t))) \right] \prod_{j \neq i} \delta_{\sigma_j, \sigma_j(t)} \] (7.23)

After averaging this expression over the random site \( i \) we obtain

\[ p_{t+\Delta}(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{2} \left[ 1 + \sigma_i \tanh(\beta h_i(\sigma(t))) \right] \prod_{j \neq i} \delta_{\sigma_j, \sigma_j(t)} \right\} \] (7.24)

If instead of the precise microscopic state \( \sigma(t) \), only the probability distribution \( p_t(\sigma') \) is given, this expression is again to be averaged over the possible states \( \sigma' \) at time \( t \)

\[ p_{t+\Delta}(\sigma) = \sum_{\sigma'} W(\sigma, \sigma') p_t(\sigma') \] (7.25)

where now

\[ W(\sigma, \sigma') = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{2} \left[ 1 + \sigma_i \tanh(\beta h_i(\sigma')) \right] \prod_{j \neq i} \delta_{\sigma_j, \sigma_j'} \right\} \] (7.26)

The transition matrix can be written in a simpler form in terms of the \( i \)-th spin-flip operator

\[ F_i \sigma = (\sigma_1, ..., \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, ..., \sigma_N) \] (7.27)

and using the definition of Kronecker deltas \( \delta_{\sigma, \sigma'} = \prod_{i=1}^{N} \delta_{\sigma_i, \sigma'_i} \)

\[ W(\sigma, \sigma') = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{2} \left[ 1 + \sigma'_i \tanh(\beta h_i(\sigma')) \right] \delta_{\sigma, \sigma'} + \frac{1}{2} \left[ 1 - \sigma'_i \tanh(\beta h_i(\sigma')) \right] \delta_{F_i \sigma, \sigma'} \right\} \] (7.28)

If we also define

\[ w_i(\sigma') = \frac{1}{2} \left[ 1 - \sigma'_i \tanh(\beta h_i(\sigma')) \right] \] (7.29)

as the probability of neuron \( i \) being updated in state \( \sigma_i = -\sigma'_i \) at time \( t + \Delta \) if it was in the opposite state \( \sigma'_i \) at time \( t \), we can write the transition matrix in yet another form

\[ W(\sigma, \sigma') = \delta_{\sigma, \sigma'} + \frac{1}{N} \sum_{i=1}^{N} \left[ w_i(\sigma') \delta_{F_i \sigma, \sigma'} - w_i(\sigma') \delta_{\sigma, \sigma'} \right] \] (7.30)
7.3. MACROSCOPIC DYNAMICS IN PROBABILISTIC FORM

Using expression (7.30) we are able to rewrite the Markov chain (7.20) as a Master equation

\[ p_{t+\Delta}(\sigma) - p_t(\sigma) = \frac{1}{N} \sum_{i=1}^{N} [w_i(F_i\sigma)p_t(F_i\sigma) - w_i(\sigma)p_t(\sigma)] \] (7.31)

In order to observe changes at a finite rate when \( N \to \infty \), we ensure that the time increment \( \Delta \) for a single sequential update has duration \( N^{-1} \). Choosing \( \Delta = N^{-1} \) and with the approximation

\[ p_{t+1/N}(\sigma) - p_t(\sigma) \approx \frac{1}{N} \frac{d}{dt} p_t(\sigma) \] (7.32)

we obtain the master equation in continuous time

\[ \frac{d}{dt} p_t(\sigma) = \sum_{i=1}^{N} [w_i(F_i\sigma)p_t(F_i\sigma) - w_i(\sigma)p_t(\sigma)] \] (7.33)

The quantities \( w_i(\sigma) \) now play the role of transition rates: in any short interval \( dt \) of rescaled time, the probability for neuron \( i \) to change state is \( w_i(\sigma)dt \).

7.3 Macroscopic dynamics in probabilistic form

In non-equilibrium statistical mechanics one aims to derive from the microscopic stochastic laws for the evolution of the system’s configuration \( \sigma \), dynamical equations for the probability distribution \( P_t(\Omega) \) of a suitable small set of macroscopic quantities \( \Omega(\sigma) = (\Omega_1(\sigma), ..., \Omega_n(\sigma)) \).

In many cases one finds, as a further simplification, that the macroscopic dynamics becomes deterministic for \( N \to \infty \). It turns out that differential equations for the probability distribution of suitably defined macroscopic state variables can be derived if the interaction matrix has a suitable structure. There will be a restriction on the number \( n \) of macroscopic state variables if we require the evolution of the variables \( \Omega \) to obey a closed set of deterministic laws.

7.3.1 Sequential dynamics

We begin to look at macroscopic dynamics for sequential update.

Toy model

We illustrate the basic ideas with the help of a simple toy model, defined by

\[ J_{ij} = \frac{J}{N} \eta_i \xi_j \quad \theta_i = 0 \] (7.34)

The local fields become \( h_i(\sigma) = J\eta_i m(\sigma) \) with \( m(\sigma) = N^{-1} \sum_k \xi_k \sigma_k \). Since they depend on \( \sigma \) only through \( m \), the latter quantity is a natural candidate observable for a macroscopic level of description. The probability of finding the value \( m(\sigma) = m \) at time \( t \) is given by

\[ P_t(m) = \sum_{\sigma} p_t(\sigma) \delta(m - m(\sigma)) \] (7.35)

Its time derivative is obtained from the continuous time master equation

\[ \frac{\partial}{\partial t} P_t(m) = \sum_{i} \sum_{\sigma} \delta(m - m(\sigma)) [w_i(F_i\sigma)p_t(F_i\sigma) - w_i(\sigma)p_t(\sigma)] \] (7.36)
CHAPTER 7. APPLICATIONS TO COMPLEX SYSTEMS

Relabeling the summation variable \( \sigma \rightarrow F_i \sigma \) in the first term

\[
\frac{\partial}{\partial t} P_t(m) = \sum_i \sum_{\sigma} w_i(\sigma)p_i(\sigma)[\delta(m - m(F_i \sigma)) - \delta(m - m(\sigma))] \tag{7.37}
\]

Because \( m(F_i \sigma) = m(\sigma) - 2\sigma_i \xi_i/N \), the arguments of the two delta functions differ only by an \( \mathcal{O}(N^{-1}) \) term, so we can make a Taylor expansion

\[
\frac{\partial}{\partial t} P_t(m) = \sum_i \sum_{\sigma} w_i(\sigma)p_i(\sigma)[\left( \frac{2}{N} \sigma_i \xi_i \right) \frac{\partial}{\partial m} \delta(m - m(\sigma)) + \mathcal{O}(N^{-2})] \tag{7.38}
\]

It may seem strange to make the Taylor expansion of a \( \delta \)-function, which is not exactly a smooth object. The validity of this operation can be confirmed by applying the expansion to an expression of the form \( \int dm P_t(m) G(m) \), where \( G \) is a smooth test function. Pulling the \( m \)-derivative in front of the sum and reordering the \( i \)-dependent terms

\[
\frac{\partial}{\partial t} P_t(m) = \frac{\partial}{\partial m} \left[ \sum_{\sigma} p_t(\sigma)\delta(m - m(\sigma)) \frac{2}{N} \sum_i \sigma_i \xi_i w_i(\sigma) \right] + \mathcal{O}(N^{-1}) \tag{7.39}
\]

and inserting (7.29)

\[
\frac{\partial}{\partial t} P_t(m) = -\frac{\partial}{\partial m} \left\{ \sum_{\sigma} p_t(\sigma)\delta(m - m(\sigma)) \right\} \times \left[ \frac{1}{N} \sum_i \xi_i \tanh(\beta h_i(\sigma) - m(\sigma)) \right] + \mathcal{O}(N^{-1}) \tag{7.40}
\]

The key simplification which allows us to eliminate \( p_t(\sigma) \) is that the local field depends on the state \( \sigma \) only through \( m(\sigma) \) so

\[
\frac{\partial}{\partial t} P_t(m) = -\frac{\partial}{\partial m} \left\{ P_t(m) \left[ \frac{1}{N} \sum_i \xi_i \tanh(\beta J m \eta_i) - m \right] \right\} + \mathcal{O}(N^{-1}) \tag{7.41}
\]

In the thermodynamic limit only the first term survives and we have a closed equation for the macroscopic probability distribution \( P_t(m) \)

\[
\frac{\partial}{\partial t} P_t(m) = -\frac{\partial}{\partial m}[P_t(m)F(m)] \tag{7.42}
\]

with

\[
F(m) = \lim_{N \to \infty} \frac{1}{N} \sum_i \xi_i \tanh(\beta J m \eta_i) - m \tag{7.43}
\]

The solution of the Liouville equation (7.42), for any function \( F(m) \), is the probability density

\[
P_t(m) = \int dm_0 P_0(m_0)\delta(m - m^*(t; m_0)) \tag{7.44}
\]

where \( m^*(t; m_0) \) denotes the solution of the differential equation

\[
\frac{d}{dt} m^* = F(m^*) \quad \text{for the initial condition } m^*(0) = m_0 \tag{7.45}
\]
7.3. MACROSCOPIC DYNAMICS IN PROBABILISTIC FORM

For $\eta_i = \xi_i$ we recover the macroscopic laws for the Hopfield model with a single pattern. In that particular case we have $F(m) = \tanh(\beta J m) - m$ and

$$\frac{dm}{dt} = \tanh(\beta J m) - m \quad (7.46)$$

For $t \to \infty$ we will converge to a point where $m = \tanh(\beta m J)$. This solution also provides a macroscopic characterization of the thermodynamic equilibrium for this model. We will see that the detailed balance condition for a stochastic dynamics to converge to equilibrium is satisfied when the interaction matrix is symmetric. In this case the long time behaviour of the system can be analyzed using equilibrium statistical mechanics. For non-symmetric interactions, on the other hand, detailed balance is absent and dynamical studies are the only option.

**General formalism**

We now generalize to less trivial choices of the interaction matrix and try to calculate the time evolution of a set of macroscopic state variables $\Omega(\sigma) = (\Omega_1(\sigma), ..., \Omega_n(\sigma))$. The equivalent of (7.47) and (7.48) are now

$$P_t(\Omega) = \sum_\sigma p_t(\sigma) \delta(\Omega - \Omega(\sigma)) \quad (7.47)$$

and

$$\frac{\partial}{\partial t} P_t(\Omega) = \sum_i \sum_\sigma \delta(\Omega - \Omega(\sigma))[w_i(F_i\sigma)p_t(F_i\sigma) - w_i(\sigma)p_t(\sigma)] \quad (7.48)$$

The difference between the arguments of the delta functions is now $\Omega_{\mu_1}^{(F_i\sigma)} - \Omega_{\mu}(\sigma) = \Delta_{ij}(\sigma)$, (this represents the sensitivity of the macroscopic observables to single neuron state change) which is no longer $O(N^{-1})$, so we cannot truncate our expansion to linear order

$$\frac{\partial}{\partial t} P_t(\Omega) = \sum_{\ell \geq 1} \frac{(-1)^\ell}{\ell!} \sum_{\mu_1=1}^n ... \sum_{\mu_\ell=1}^n \delta(\Omega - \Omega(\sigma))[P_t(\Omega)F^{(\ell)}_{\mu_1,...,\mu_\ell}(\Omega; t)] \quad (7.49)$$

with

$$F^{(\ell)}_{\mu_1,...,\mu_\ell}(\Omega; t) = \left( \sum_{j=1}^N w_j(\sigma) \Delta_{j\mu_1}(\sigma) ... \Delta_{j\mu_\ell}(\sigma) \right)_{\Omega,t} \quad (7.50)$$

defined in terms of the conditional (or sub-shell) average

$$\langle ... \rangle_{\Omega,t} = \frac{\sum_\sigma p_t(\sigma) \delta(\Omega - \Omega(\sigma))}{\sum_\sigma p_t(\sigma) \delta(\Omega - \Omega(\sigma))} \quad (7.51)$$

As in the previous toy model, the Kramers-Moyal expansion (7.49) is to be interpreted in a distributional sense, that is, only to be used in expressions of the form $\int d\Omega P_t(\Omega)G(\Omega)$, with sufficiently smooth function $G(\Omega)$. The first $(\ell = 1)$ term gives the Liouville equation which describes deterministic flow in the $\Omega$ space, driven by the flow field $F^{(1)}$. Including the second term $(\ell = 2)$ gives the Fokker-Planck equation which in addition to the flow also describes diffusion of the macroscopic probability density $P_t(\Omega)$, generated by the diffusion matrix $\{F^{(2)}_{\mu\nu}\}$. A sufficient condition for the observables $\Omega(\sigma)$ to evolve in time deterministically in the limit $N \to \infty$ can be derived by using the distributional interpretation of the Kramers-Moyal expansion and making the derivatives $\partial/\partial \Omega_{\mu}$ apply to the function $G$ alone, via integration.
by parts; since the latter are (uniformly) bounded by assumption, and $|w_j(\sigma)| < 1$, the condition to have the $\ell \geq 2$ terms in the expansion suppressed as $N \to \infty$ is

$$ \lim_{N \to \infty} \sum_{\ell \geq 2} \frac{1}{\ell!} \sum_{\mu_1=1}^{n} \ldots \sum_{\mu_\ell=1}^{n} w_j(|\Delta_{j\mu_1}(\sigma)\ldots\Delta_{j\mu_\ell}(\sigma)|) \Omega_a = 0 $$

(7.52)

If for a given set of macroscopic observables this condition is satisfied (this will impose restriction on the number $n$) we can for large $N$ describe the evolution of the macroscopic probability distribution by the Liouville equation

$$ \frac{\partial}{\partial t} P_t(\Omega) = - \sum_{\mu=1}^{n} \frac{\partial}{\partial \Omega_{\mu}} [ P_t(\Omega) F^{(1)}(\Omega) ] $$

(7.53)

the solution of which describes deterministic flow

$$ P_t(\Omega) = \int d\Omega_0 P_0(\Omega_0) \delta(\Omega - \Omega^*(t; \Omega_0)) $$

$$ \frac{d}{dt} \Omega^* = F^{(1)}(\Omega^*(t; \Omega_0); t) \quad \Omega^*(0; \Omega_0) = \Omega_0 $$

(7.54)

Equations (7.54) will in general not be autonomous. A way to allow the elimination of the explicit time dependence and thereby turn the state variable $\Omega$ into an autonomous level of description is to choose (if possible) the macroscopic state variables $\Omega$ in such a way that there is no explicit time dependence in the flow field $F^{(1)}(\Omega; t)$, i.e. there must exist a vector field $\Phi(\Omega)$ such that

$$ \lim_{N \to \infty} \sum_{j=1}^{N} w_j \Delta_j(\sigma) = \Phi(\Omega(\sigma)) $$

(7.55)

If for example, one chooses the macroscopic state variable $\Omega_{\mu}$ to depend linearly on the microscopic state variables $\sigma$, i.e.

$$ \Omega(\sigma) = \frac{1}{N} \sum_{j=1}^{N} \omega_j \sigma_j $$

(7.56)

then we obtain

$$ \lim_{N \to \infty} \sum_{j=1}^{N} w_j \Delta_j(\sigma) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \omega_j \tanh(\beta h_j(\sigma)) - \Omega $$

(7.57)

Then, the only further condition to eliminate the explicit time dependence from $F^{(1)}$ is that all local fields $h_k$ depend (to leading order in $N$) on the microscopic state $\sigma$ only through the macroscopic state variables $\Omega$. Since the local fields depend linearly on $\sigma$ this in turn implies that the interaction matrix must be separable, i.e. $J_{ij} = Q(\xi_i, \xi_j) = N^{-1} \xi_i A \xi_j$. The choice $A_{\mu,\nu} = \delta_{\mu,\nu}$ corresponds to the Hopfield model with simultaneous storage of several patterns, $\xi^\mu = (\xi^\mu_1, \ldots, \xi^\mu_N)$, ($\mu = 1, \ldots, p$). Taking then $\omega^\mu_j = \xi^\mu_j$, $\Omega_{\mu}$ becomes the overlap between the $\mu$-th stored pattern and the state of the system. The fixed point equations can be written then as

$$ \Omega_{\mu} = \langle \xi^\mu \tanh(\beta \sum_{\nu} \xi^\nu \Omega_{\nu}) \rangle_{\xi^\mu} $$

(7.58)

where $\langle f(\xi) \rangle_{\xi} = N^{-1} \sum_i f(\xi_i)$. If it is not possible to find a set of macroscopic observables that satisfies (7.55) additional assumptions or restrictions are needed.
7.3. MACROSCOPIC DYNAMICS IN PROBABILISTIC FORM

7.3.2 Parallel dynamics

We now turn to the case of parallel dynamics, i.e. to the discrete time stochastic microscopic laws (7.20) and try to get discrete mappings (instead of differential equations) for the evolution of macroscopic probability distribution.

Toy model

We first see what happens to our toy model. The evolution of the macroscopic probability distribution is easily obtained as

\[ P_{t+1}(m) = \sum_{\sigma'} \delta(m - m(\sigma'))W(\sigma, \sigma')p_t(\sigma') \]

\[ = \int dm' \tilde{W}_t(m, m')P_t(m') \]  \hspace{1cm} (7.59)

with

\[ \tilde{W}_t(m, m') = \frac{\sum_{\sigma} \delta(m - m(\sigma))\delta(m' - m(\sigma'))W(\sigma, \sigma')p_t(\sigma')}{\sum_{\sigma'} \delta(m' - m(\sigma'))p_t(\sigma')} \]  \hspace{1cm} (7.60)

We next insert

\[ W(\sigma, \sigma') = \prod_i \frac{e^{\beta h_i(\sigma')\xi_i}}{2 \cosh(\beta h_i(\sigma'))}, \quad h_i(\sigma) = J\eta_i m(\sigma) \]  \hspace{1cm} (7.61)

and \( m(\sigma) = N^{-1} \sum_k \xi_k \sigma_k \), and we try to describe the dynamics at the macroscopic level of \( m(\sigma) \). Because the fields only depend on the microscopic state \( \sigma \) through \( m(\sigma), p_t(\sigma) \) drops out of \( \tilde{W}_t \) and this loses its explicit time dependence, i.e. \( \tilde{W}_t(m, m') \rightarrow \tilde{W}(m, m') \):

\[ \tilde{W}(m, m') = e^{-\sum_i \ln \cosh(\beta Jm'_i \eta_i)} \langle \delta[m - m(\sigma)] e^{\beta J \sum_i \eta_i \xi_i} \rangle_{\sigma} \]  \hspace{1cm} (7.62)

with \( \langle \ldots \rangle_{\sigma} = 2^{-N} \sum_{\sigma} \ldots \). Inserting the integral representation for the \( \delta \)-function

\[ \delta(m - m(\sigma)) = \frac{\beta N}{2\pi} \int dk e^{i\beta km - i\beta \sum_i \xi_i \sigma_i} \]  \hspace{1cm} (7.63)

allows to preform the spin average, giving

\[ \tilde{W}(m, m') = \frac{\beta N}{2\pi} \int dk e^{N \psi(m, m', k)} \]  \hspace{1cm} (7.64)

with

\[ \psi(m, m', k) = i\beta km + \langle \ln \cosh(\beta Jm'_i - i\beta k \xi_i) \rangle_{\eta_i} - \langle \ln \cosh(\beta Jm'_i) \rangle_{\eta} \]  \hspace{1cm} (7.65)

where \( \langle f(\eta) \rangle_{\eta} = N^{-1} \sum_i f(\eta_i) \). Since the kernel \( \tilde{W}(m, m') \) is normalized by construction, i.e. \( \int dm \tilde{W}(m, m') = 1 \), we find that for \( N \rightarrow \infty \) the expectation value with respect to \( \tilde{W}(m, m') \) of any sufficiently smooth function \( f(m) \) will be determined only by the value \( m^*(m') \) of \( m \) in the relevant saddle point of \( \psi \):

\[ \int dm f(m) \tilde{W}(m, m') = \frac{\int dmdk f(m)e^{N \psi(m, m', k)}}{\int dmdk e^{N \psi(m, m', k)}} \rightarrow f(m^*(m')) \]  \hspace{1cm} (N \rightarrow \infty)  \hspace{1cm} (7.66)

Variation of \( \psi \) with respect to \( k \) and \( m \) gives the saddle point equation

\[ m = \langle \xi \tanh(\beta Jm' - \beta k \xi) \rangle_{\eta \xi}, \quad k = 0 \]  \hspace{1cm} (7.67)
We can now conclude from (7.66) that
\[
\lim_{N \to \infty} \tilde{W}(m, m') = \delta(m - m^*(m')) \quad m^*(m') = \langle \xi \tanh(\beta J \eta m') \rangle_{\eta, \xi}
\] (7.68)
and the macroscopic equation becomes
\[
P_{t+1}(m) = \int dm' \delta[m - \langle \xi \tanh(\beta J \eta m' - \beta \xi k) \rangle_{\eta, \xi}] P_t(m')
\] (7.69)
which describes deterministic evolution. If at time 0 we know \(m\) exactly, this will remain so for finite timescale, and \(m\) will evolve according to a discrete version of the sequential dynamical law (7.46)
\[
m_{t+1} = \langle \xi \tanh(\beta J \eta m_t) \rangle_{\eta, \xi}
\] (7.70)

### 7.3.3 General formalism

For more general sets of intensive macroscopic variables \(\Omega(\sigma) = (\Omega_1(\sigma), ..., \Omega_n(\sigma))\) an equation for
\[
P_t(\Omega) = \sum_\sigma p_t(\sigma) \delta(\Omega - \Omega(\sigma))
\] (7.71)
is obtained as
\[
P_{t+1}(\Omega) = \int d\Omega' \tilde{W}_t(\Omega, \Omega') P_t(\Omega')
\] (7.72)
with
\[
\tilde{W}(\Omega, \Omega') = \sum_{\sigma \sigma'} \frac{\delta(\Omega - \Omega(\sigma)) \delta(\Omega' - \Omega(\sigma')) W(\sigma, \sigma') p_t(\sigma')}{\sum_{\sigma'} \delta(\Omega' - \Omega(\sigma')) p_t(\sigma')}
\] (7.73)
and
\[
W(\sigma, \sigma') = \prod_i \frac{e^{\beta h_i(\sigma)}}{2 \cosh(\beta h_i(\sigma'))}
\] (7.74)
If the local fields \(h_i(\sigma)\) depend on the microscopic state \(\sigma\) only through the values of \(\Omega(\sigma)\), again \(\tilde{W}_t(\Omega, \Omega') \to \tilde{W}(\Omega, \Omega')\) and
\[
\tilde{W}(\Omega, \Omega') = \left(\frac{\beta N}{2 \pi}\right)^n \int dK e^{N \psi(\Omega, \Omega', K)}
\] (7.75)
with
\[
\psi(\Omega, \Omega', K) = i\beta K \Omega + \frac{1}{N} \left(e^{\beta \sum_i h_i(\Omega') - i N K \Omega(\sigma)} - 1\right) \sigma - \frac{1}{N} \sum_i \ln \cosh(\beta h_i(\Omega'))
\] (7.76)
Using the normalization relation \(\int d\Omega \tilde{W}(\Omega, \Omega') = 1\), we can write expectation values with respect to \(\tilde{W}(\Omega, \Omega')\) of macroscopic quantities \(f(\Omega)\) as
\[
\int d\Omega f(\Omega) \tilde{W}(\Omega, \Omega') = \frac{\int d\Omega dk f(\Omega) e^{N \psi(\Omega, \Omega', k)}}{\int d\Omega dk e^{N \psi(\Omega, \Omega', k)}} \to f(\Omega^*)(\Omega') \quad (N \to \infty)
\] (7.77)
For the saddle point method to apply in determining the leading order in \(N\) of the average (7.77) we encounter a restriction on the number \(n\) of our macroscopic variables, since \(n\) determines the dimension of the integration concerned. (We must require that Gaussian fluctuations around
the maximum of $\psi$ are suppressed as $N \to \infty$, which is only guaranteed if $\lim_{N \to \infty} n/\sqrt{N} = 0$). Variation of $\psi$ with respect to $K$ and $\Omega$ gives the saddle point equation

$$\Omega = \frac{\langle \Omega(\sigma)e^{\beta\sum_i \sigma_i h_i(\Omega)}\rangle_{\sigma}}{\langle e^{\beta\sum_i \sigma_i h_i(\Omega)}\rangle_{\sigma}}; \quad K = 0$$ (7.78)

We can now conclude that

$$\lim_{N \to \infty} \tilde{W}(\Omega, \Omega') = \delta(\Omega - \Omega^*(\Omega')) \quad \Omega^*(\Omega') = \frac{\langle \Omega(\sigma)e^{\beta\sum_i \sigma_i h_i(\Omega')}\rangle_{\sigma}}{\langle e^{\beta\sum_i \sigma_i h_i(\Omega')}\rangle_{\sigma}}$$ (7.79)

and the macroscopic equation becomes

$$P_{t+1}(\Omega) = \int d\Omega' \delta \left( \Omega - \frac{\langle \Omega(\sigma)e^{\beta\sum_i \sigma_i h_i(\Omega')}\rangle_{\sigma}}{\langle e^{\beta\sum_i \sigma_i h_i(\Omega')}\rangle_{\sigma}} \right) P_t(\Omega')$$ (7.80)

This equation again describes deterministic evolution

$$\Omega_{t+1} = \frac{\langle \Omega(\sigma)e^{\beta\sum_i \sigma_i h_i(\Omega(t))}\rangle_{\sigma}}{\langle e^{\beta\sum_i \sigma_i h_i(\Omega(t))}\rangle_{\sigma}}$$ (7.81)

Exercise Consider a simple system, consisting of $N = 2$ neurons with no self-interactions and no external inputs, so that the fields acting on neuron 1 and 2 are, respectively

$$h_1(\sigma) = J_{12} \sigma_2 \quad h_2(\sigma) = J_{21} \sigma_1$$ (7.82)

where the vector $\sigma = (\sigma_1, \sigma_2)$ collects the four possible states of the two neurons. Write down the transition probabilities $W(\sigma, \sigma')$ for both parallel and sequential dynamics and verify that is satisfies the properties of a stochastic matrix.

7.4 Detailed balance

We have shown that the dynamics of a recurrent network of binary units takes the form of a Markov chain, for both parallel and sequential updates

$$p_{t+1}(\sigma) = \sum_{\sigma'} W(\sigma', \sigma') p_t(\sigma)$$ (7.83)

From the theory of Markov chains, we know that if the transition matrix is regular, then the Markov chain is ergodic and the probability distribution $p_t(\sigma)$ of a finite system is guaranteed to converge to a unique stationary distribution $p_\infty(\sigma)$ for $t \to \infty$. If, in addition, the system satisfies detailed balance, the resulting stationary state is called an equilibrium state and $p_\infty(\sigma)$ has the form of a Boltzmann distribution.
7.4.1 Interaction symmetry

The unique stationary distribution \( p_\infty(\sigma) \) over the \( 2^N \) possible microscopic states \( \sigma \), which is reached in the limit \( N \to \infty \), is determined by the condition

\[
\text{for all } \sigma \in \{-1, 1\}^N: \quad p_\infty(\sigma) = \sum_{\sigma'} W(\sigma, \sigma') p_\infty(\sigma') \tag{7.84}
\]

While it is reassuring that \( p_\infty \) exists, to calculate it we would have to solve this system of \( 2^N \) linear equation for the \( 2^N \) values \( p_\infty(\sigma) \), subject to conditions \( p_\infty(\sigma) \geq 1 \) and \( \sum_\sigma p_\infty(\sigma) = 1 \). This seems rather a hopeless task.

Detailed balance is a special feature which greatly simplifies the calculation of the stationary distribution. For systems which satisfy detailed balance the stationary distribution is such that

\[
W(\sigma, \sigma') p_\infty(\sigma') = W(\sigma', \sigma) p_\infty(\sigma) \tag{7.85}
\]

We note that any distribution which satisfies (7.85) is necessarily stationary; the converse is not true.

Next we show that for our neural network models, detailed balance holds if and only if the interactions are symmetric, i.e. \( J_{ij} = J_{ji} \) for all neuron pairs \((i, j)\) (for sequential dynamics there must also be no self-interactions). We consider the case of parallel and sequential dynamics separately.

**Parallel dynamics**

For parallel dynamics, the detailed balance condition becomes

\[
e^\beta \frac{\sum_{i=1}^N \sigma_i h_i(\sigma') p_\infty(\sigma')}{\prod_{i=1}^N \cosh(\beta h_i(\sigma'))} = e^\beta \frac{\sum_{i=1}^N \sigma' i h_i(\sigma) p_\infty(\sigma)}{\prod_{i=1}^N \cosh(\beta h_i(\sigma))} \tag{7.86}
\]

with \( h_i(\sigma) = \sum_j J_{ij} \sigma_j + \theta_i \) The transition matrix indeed describes an ergodic system. Since from any initial state one can reach any final state with nonzero probability, the stationary probability \( p_\infty(\sigma) \) must be nonzero. Without loss of generality we can therefore put

\[
p_\infty(\sigma) = e^{\beta \sum_{i=1}^N \theta_i \sigma_i + K(\sigma)} \prod_{i=1}^N \cosh(\beta h_i(\sigma)) \tag{7.87}
\]

Detailed balance then simplifies to

\[
K(\sigma) - K(\sigma') = \sum_{ij} \sigma_i (J_{ij} - J_{ji}) \sigma'_j \tag{7.88}
\]

**Exercise** Show that if interactions are symmetric this conditions is satisfied by \( K(\sigma) = K \) (\( K \) a constant, which is determined from normalization of \( p_\infty(\sigma) \)) and the equilibrium distribution has the Boltzmann form

\[
p_\infty(\sigma) = \frac{1}{Z} e^{\beta H(\sigma)} \quad Z = \sum_\sigma e^{-\beta H(\sigma)} \tag{7.89}
\]
7.4. DETAILED BALANCE

with the pseudo-Hamiltonian (the Hamiltonian still depends on the noise parameter $\beta$)

$$H(\sigma) = -\sum_i \theta_i \sigma_i - \frac{1}{\beta} \ln \cosh(\beta h_i)$$  \hspace{1cm} (7.90)

Conversely, if detailed balance holds, i.e. there is a function $K(\sigma)$ obeying (7.88), taking an average over the $2^N$ states $\sigma'$ of (7.88) shows that $K(\sigma) = \langle K(\sigma') \rangle \forall \sigma$, so $K(\sigma)$ is again constant, and this implies symmetry of interactions.

Sequential dynamics without self-interaction

Now consider the case of sequential dynamics without self-interactions, i.e. $J_{ii} = 0$. Detailed balance is now non-trivial only if $\sigma' = F_i \sigma$, where (7.85) simplifies to

$$e^{-\beta(F_i \sigma_i) h_i(F_i \sigma)} p_\infty(F_i \sigma) = e^{-\beta \sigma_i h_i(\sigma)} p_\infty(\sigma) \cosh(\beta h_i(\sigma)) \quad \text{for all } \sigma \text{ and all } i$$  \hspace{1cm} (7.91)

with $h_i(\sigma) = \sum_{j \neq i} J_{ij} \sigma_j + \theta_i$. Note that $h_i(\sigma) = h_i(F_i \sigma)$.

Since the system is ergodic, all stationay probabilities $p_\infty(\sigma)$ must be non-zero and we can write

$$p_\infty(\sigma) = e^{\beta \sum_k \theta_k \sigma_k + \frac{1}{2} \sum_{k \neq \ell} \sigma_k J_{k\ell} \sigma_\ell + K(\sigma)}$$  \hspace{1cm} (7.92)

Exercise Show that the detailed balance condition simplifies to

$$K(F_i \sigma) - K(\sigma) = \sum_{k \neq i} \sigma_i (J_{ki} - J_{ik}) \sigma_k \sigma_i$$  \hspace{1cm} (7.93)

and that if the interactions are symmetric, this is satisfied by $K(\sigma) = K$ ($K$ a constant, determined from normalization of $p_\infty$). The stationary probability distribution then takes the Boltzmann form

$$p_\infty(\sigma) = \frac{1}{Z} e^{\beta H(\sigma)} \quad Z = \sum_\sigma e^{-\beta H(\sigma)}$$  \hspace{1cm} (7.94)

with the Hamiltonian

$$H(\sigma) = -\frac{1}{2} \sum_{k \neq \ell} J_{k\ell} \sigma_k \sigma_\ell - \sum_k \theta_k \sigma_k$$  \hspace{1cm} (7.95)

Conversely, if detailed balance holds, i.e. there is a function $K(\sigma)$ obeying (7.93),

$$K(F_j F_i \sigma) - K(F_j \sigma) - [K(F_i \sigma) - K(\sigma)] = -2 \sigma_i (J_{ji} - J_{ij}) \sigma_j$$  \hspace{1cm} (7.96)

Since the LHS is symmetric under permutation of the pair $(i, j)$ so must be the RHS, which implies symmetry of interactions, i.e. $J_{ij} = J_{ji}$. Taking an average over the $2^N$ states $\sigma'$ of (7.88) shows that $K(\sigma) = \langle K(\sigma') \rangle \forall \sigma$, so $K(\sigma)$ is again constant, and this implies symmetry of interactions.
Chapter 8

Generating functional analysis

8.1 General

The approach followed in the previous chapter to derive closed macroscopic laws from the microscopic equations fails when the number of attractors of the dynamics is no longer small compared to the number $N$ of microscopic neuronal variables. This is due to the presence of a number of "disorder" variables per degree of freedom which is proportional to $N$, over which we are forced to average the macroscopic laws. This situation is reflected in the inability to find an exact set of closed equations for a finite number of observables. The only fully exact procedure available at present is generating functional analysis, also called "path integral formalism" or "dynamic mean-field theory" and is based on a philosophy different from the one described so far. Rather than working with the probability $p_t(\sigma)$ of finding a microscopic state $\sigma$ at time $t$ in order to calculate the statistics of a set of macroscopic observables $\Omega(\sigma)$ at time $t$, one here turns to the probability $\text{Prob}[\sigma(0),...,\sigma(t_m)]$ of finding a microscopic path $\sigma(0) \rightarrow \sigma(t_1) \rightarrow ... \rightarrow \sigma(t_m)$. The idea is to concentrate on the moment generating function $Z$, which like $\text{Prob}[\sigma(0),...,\sigma(t_m)]$ fully captures the statistics of paths. Correlations and response functions will be the language in which the generating functional methods are formulated. We first illustrate how to calculate the generating function for a system described by one degree of freedom and then we move to consider systems whose state is described by a vector.

8.2 Generating function

Consider a one-dimensional random variable $\phi$ evolving according to the following Markov process:

$\phi(m+1) = f(\phi(m)) + \eta(m)$

(8.1)

where $\eta(m)$ is a random variable generated independently at each time step from a Gaussian probability density $p(\eta)$ with zero mean and variance $2T$. The "temperature" $T$ govern the size of the fluctuations. Instead of studying the evolution of a single instance of $\phi$ one can study the evolution of the probability distribution $p_m(\phi)$. This evolution is described by

$p_{m+1}(\phi) = \int d\phi' W_m(\phi|\phi')p_m(\phi')$

(8.2)
where the transition probability is given by

$$W_m(\phi|\phi') = \int d\eta p(\eta) \delta(\phi - f(\phi') - \eta)$$

We introduce the following representation of the Dirac delta-function

$$\delta(\phi) = \int \sqrt{2\pi} e^{i\hat{\phi}\phi}$$

We integrate over the Gaussian noise \(\eta\) by using the identity

$$\int dx \sqrt{2\pi} e^{-\frac{1}{2}ax^2 - bx} = e^{-\frac{b^2}{4a}}$$

thus obtaining

$$W_m(\phi|\phi') = \int d\hat{\phi} \exp \left[ i\hat{\phi}(\phi - f(\phi')) - T\hat{\phi}^2 \right]$$

We study the probability that a certain path in time is taken by \(\phi\). The probability for a particular path from step \(m = 0\) to step \(m = M\) can be written as

$$p(\phi(M), ..., \phi(1), \phi(0)) = \prod_{m=0}^{M-1} W_m(\phi(m + 1)|\phi(m)) p_0(\phi(0))$$

We will denote averages over paths as \(\langle ... \rangle\). A convenient way to study the joint probability distribution (8.7) is by its characteristic or moment generating function

$$Z_M[\psi] = \langle \exp \left[ \sum_{m=0}^{M} i\psi(m)\phi(m) \right] \rangle = \int \prod_{m=0}^{M} d\phi(m) p(\phi(M), ..., \phi(0)) \exp[\sum_{m=0}^{M} i\psi(m)\phi(m)]$$

This function, which is like the Fourier transform of the path probability distribution, can be used to generate all moments of the distribution. For example, the first and the second moments ("magnetization" and two-time correlation) are given by

$$M(m) \equiv \langle \phi(m) \rangle = \frac{\partial Z_M[\psi]}{\partial i\psi(m)}|_{\psi=0}$$

$$C(m,n) \equiv \langle \phi(m)\phi(n) \rangle = \frac{\partial^2 Z}{\partial \psi(m) \partial \psi(n)}|_{\psi=0}$$

Upon introducing the notation \(D\phi = \prod_{m=0}^{M-1} [d\phi/\sqrt{2\pi}]\) for path integration, the full generating function looks like

$$Z_M[\psi] = \int d\phi(M) D\phi D\hat{\phi} \exp \left[ S[\phi, \hat{\phi}] + \sum_{m=0}^{M} i\psi(m)\phi(m) \right] p_0(\phi(0))$$

where in the effective action

$$S[\phi, \hat{\phi}] = \sum_{m=0}^{M-1} i\hat{\phi}(m)[\phi(m+1) - f(\phi(m))] - \frac{1}{2} \sum_{m=0}^{M-1} 2T\hat{\phi}^2(m)$$
only appears the equation of motion without noise and the effect of having a noise is expressed only through the noise variance $2T$. When we introduce a time dependent field in the system, i.e.

$$\phi(m + 1) = f(\phi(m)) + \theta(m) + \eta(m) \quad (8.12)$$

we will have in $S$ an extra term $\sum_{m=0}^{M-1} \hat{\phi}(m)\theta(m)$. So if we derive the average of any quantity with respect to $\theta(n)$ we will pull down a factor $\hat{\phi}(n)$ from the exponential. The response of $M(m)$ to an infinitesimally small external field $\theta$ at step $n$ is

$$G(m,n) \equiv \frac{\partial M(m)}{\partial \theta(n)} = \frac{\partial^2 Z}{\partial \theta(n) \partial \psi(m)}|_{\psi=0} = \langle i\hat{\phi}(n)\phi(m) \rangle (8.13)$$

Because of the normalization of the generating function $Z[0] = 1$, averages involving only $\hat{\phi}$ are zero, for example

$$\langle \hat{\phi}(m) \rangle = \frac{\partial Z[0]}{i \partial \theta(m)} = 0, \quad \langle \hat{\phi}(m)\hat{\phi}(n) \rangle = \frac{\partial^2 Z}{i \partial \theta(m) \partial \theta(n)} = 0 \quad (8.14)$$

These rather simple identities may be helpful in reducing the complexity of the problems later on.

## 8.3 Langevin dynamics

We now consider the dynamics of one degree of freedom $\phi(t)$, which evolves according to the Hamiltonian $H[\phi(t)]$, in presence of noise $\eta(t)$. The Langevin equation is

$$\dot{\phi} = \frac{\partial H[\phi(t)]}{\partial \phi(t)} + \eta(t) \quad (8.15)$$

where $\eta$ is a Gaussian white noise with properties:

$$\langle \eta(t) \rangle = 0, \quad \langle \eta(t)\eta(t') \rangle = 2T\delta(t-t'). \quad (8.16)$$

If we want to construct a generating function for this continuous time process, it is natural to discretise time in steps of size $\Delta$. The evolution equation gives

$$\phi(t + \Delta) = \phi(t) + \int_t^{t+\Delta} dt' \frac{\partial H[\phi(t')]}{\partial \phi(t')} + \int_t^{t+\Delta} dt' \eta(t') \quad (8.17)$$

At this point we have a choice in how to calculate the integral of the force $f(\phi(t')) = \frac{\partial H[\phi(t')]}{\partial \phi(t')}$. The simplest way is to apply the Ito interpretation to the Langevin equation and introduce a discrete time $t = m\Delta$ and discretize the equation in the following way

$$\phi(m + 1) = \phi(m) + \Delta f(\phi(m)) + \tilde{\eta}(m) \quad (8.18)$$

where the integrated noise $\tilde{\eta}(m) = \int_{m\Delta}^{(m+1)\Delta} dt' \eta(t')$ is Gaussian with zero mean and variance

$$\langle \tilde{\eta}(m)\tilde{\eta}(n) \rangle = 2T\Delta \delta_{m,n} \quad (8.19)$$
Up to some factors $\Delta$ this process is identical to the one in the previous section, so we can easily write down the generating function corresponding to our new process (we only rescale $\psi(m) \to \Delta \psi(m)$):

$$Z_M[\psi] = \int d\phi(M)D\phi D\hat{\phi} \exp \left[ \Delta S[\phi, \hat{\phi}] + \Delta \sum_{m=0}^{M} i\psi(m)\phi(m) \right] p_0(\phi(0)) \tag{8.20}$$

where

$$S[\phi, \hat{\phi}] = \sum_{m=0}^{M-1} i\hat{\phi}(m)\left[\frac{\phi(m+1) - \phi(m)}{\Delta} - f(\phi(m))\right] - \frac{1}{2} \sum_{m=0}^{M-1} 2T\dot{\phi}^2(m) \tag{8.21}$$

At this point we can take the limit $\Delta \to 0$ to recover a description of the original process. This replaces $\Delta \sum_m$ by $\int dt'$. The generating function becomes a generating functional on the function $\psi(t)$, which can be written as

$$Z_M[\psi] = \int D\phi D\hat{\phi} \exp \left[ S[\phi, \hat{\phi}] + \int_0^\infty dt' i\psi(t')\phi(t) \right] p_0(\phi(0)) \tag{8.22}$$

where the effective action is

$$S[\phi, \hat{\phi}] = \int_0^\infty dt' i\hat{\phi}(t')\left[\phi(t') - f(\phi(t))\right] - \frac{1}{2} \int_0^\infty dt' 2T\dot{\phi}^2(t) \tag{8.23}$$

and we have introduced $\hat{\phi}(t) = \lim_{\Delta \to 0}(\phi(t + \Delta) - \phi(t)/\Delta)$. A rigorous definition of the continuous time limit of the integrals over $D\phi$ and $D\hat{\phi}$ has been given by Wiener. For the time being we ignore this difficulty and treat the integrals over the paths like a product of many normal integrals. An alternative derivation of (8.22) can be obtained by starting with the (discretised version of the) probability measure on a path $\eta(t)$. We assume $\eta$ is a Gaussian white noise, i.e. drawn from the distribution

$$p[\eta] \propto \exp \left[ -\frac{1}{2} \int dt dt' \eta(t) D^{-1}(t-t')\eta(t') \right]$$

with $\langle \eta(t)\eta(t') \rangle = 2T\delta(t-t') = D(t-t')$ and $\langle \eta(t) \rangle = 0. \tag{8.24}$

The distribution $p[\eta]$ induces a distribution of $x$, $P[x]$.

In order to calculate the average over the noise of a generic observable $A$, function of the degree of freedom $x(t)$, one performs

$$\langle A(x) \rangle = \int D\eta p(\eta) A(x,\eta) = \int D\eta p(\eta) \int dx \delta(x-x_\eta) A(x)$$

$$= \int D\eta p(\eta) \int dx \delta(E[x(t)]) \left| \frac{\delta E}{\delta x} \right| A(x) \tag{8.25}$$

where $E[x(t)] = \partial_t x + \partial_x H - \eta$. In the last line we used the identity

$$\delta[f(x)] = \frac{\delta(x-x_0)}{|f'(x)|} \quad \text{with} \quad f(x_0) = 0 \tag{8.26}$$

On the other hand, $\langle A(x) \rangle = \int dx P(x) A(x)$, so the noise-induced distribution of $x$ is determined as

$$P(x) = \int D\eta p(\eta) \delta(E[x(t)]) \left| \frac{\delta E}{\delta x} \right| \tag{8.27}$$
8.4. A SIMPLE EXAMPLE: THE HARMONIC OSCILLATOR

In order to calculate the determinant of the Jacobian

\[ J(t, t') = \left| \frac{\delta E(x(t))}{\delta x(t')} \right| \]

of the transformation \( x \rightarrow E(x) \), we rewrite the Langevin equation in integral form

\[ x(t + \tau) = x(t) - \int_t^{t+\tau} dt' \frac{\partial H(x(t'))}{\partial x} + \int_t^{t+\tau} dt' \eta(t') \]

The integral of the force \( F(x(t')) = \frac{\partial H(x(t'))}{\partial x} \) can be approximated in two ways as \( \tau \rightarrow 0 \):

\[ \int_t^{t+\tau} dt' F(x(t')) \approx \begin{cases} \tau F(x(t)) & \text{Ito prescription} \\ \tau F(x(t)) + F(x(t + \tau))/2 & \text{Stratonovich prescription} \end{cases} \]

We use Ito prescription and discretize the equation

\[ x_k = x_{k-1} - \tau F(x_{k-1}) + W_k \]

The element \( J_{kl} \) of the Jacobian can now be found by calculating the variation of

\[ E_k = x_k - x_{k-1} + \tau F(x_{k-1}) - W_k \]

when \( x_l \) is changed

\[ J_{kl} = \frac{\partial E_k}{\partial x_l} = \begin{cases} 1 & k = l \\ 1 + \tau F' & l = k - 1 \\ 0 & \text{otherwise} \end{cases} \rightarrow J = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ y & 1 & 0 & \cdots & \vdots \\ 0 & y & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & y & 1 \end{pmatrix} \]

so the determinant of the Jacobian is \(|J| = 1\), and going back to continuous time we have

\[ P(x) = \int D\eta p(\eta) \delta(E[x(t)]) \]

If we use Stratonovich definition \( J \neq 1 \) and the equation gains a term which is not present in the Ito generating functional. As long as we have additive noise in the Langevin equation, all measurable quantities are unaffected by the choice of the interpretation of the Langevin equation. In the case of multiplicative noise, i.e. if we consider noise terms like \( g(\phi(t))\eta(t) \) the choice to evaluate \( g(\phi(t)) \) before changing \( \phi \) due to the impact of the noise term at time \( t \) or to approximate the noise by a smoother function has an influence on the evolution.

8.4 A simple example: the harmonic oscillator

We consider the evolution of one degree of freedom system, whose dynamics, in absence of noise, is governed by the Hamiltonian

\[ H = \frac{1}{2} m x^2 \]
In presence of noise the time evolution of the system is given by the Langevin equation
\[ \partial_t x = -mx + \eta(t) \] (8.34)
and we assume Gaussian white noise, i.e.
\[ \langle \eta(t) \rangle = 0 \quad \langle \eta(t) \eta(t') \rangle = D(t-t') \] (8.35)
The generating functional is given by
\[ Z = \int DxD\hat{x} \exp \left\{ -\frac{1}{2} \int dt dt' A(t,t') \hat{x}(t)\hat{x}(t') + i \int dt [\hat{x}(t)\partial_t x + \hat{x}(t)mx(t)] \right\} \] (8.36)
Because \( i\hat{x}\partial_t x = -ix\partial_t \hat{x} \), we can rewrite the first term in the square brackets as
\[ \hat{x}\partial_t x = \frac{1}{2} \hat{x}\partial_t x - \frac{1}{2} x\partial_t \hat{x} \]
So if we define
\[ \phi(t) = (\hat{x}(t), x(t)) \] (8.37)
the generating functional can be expressed as
\[ Z = \int D\phi \exp \left\{ -\frac{1}{2} \int dt dt' A(t,t') \phi(t)\phi(t') \right\} \] (8.38)
where
\[ A(t,t') = \begin{pmatrix} D & -i\partial_t - im \\ i\partial_t - im & 0 \end{pmatrix} \]
The elements of the inverse matrix
\[ A^{-1} = \begin{pmatrix} 0 & GR \\ GA & C \end{pmatrix} \] (8.39)
will give the propagators\(^1\) we are interested in (we first state the results and then we prove them):
\[ \langle x(t) x(t') \rangle = [A^{-1}(t,t')]_{22} \] (8.40)
\[ \langle \hat{x}(t) \hat{x}(t') \rangle = [A^{-1}(t,t')]_{11} = 0 \] (8.41)
\[ \langle x(t) \hat{x}(t') \rangle = [A^{-1}(t,t')]_{21} = GA(t,t') \] advanced propagator (8.42)
\[ \langle \hat{x}(t) x(t') \rangle = [A^{-1}(t,t')]_{12} = GR(t,t') \] retarded propagator (8.43)
with
\[ GA(\omega) = \frac{1}{\omega - im} \quad GR(\omega) = -\frac{1}{\omega + im} \quad C(\omega) = \frac{D(\omega)}{\omega^2 + m^2} \] (8.44)
We first prove the results stated in (8.40,8.41,8.42,8.43) and then we prove (8.44). For the correlator one has
\[ \langle xx \rangle = \frac{\int DxD\hat{x}xe^{-\frac{1}{2}\phi(t)A(t,t')\phi(t')}}{\int DxD\hat{x}e^{-\frac{1}{2}\phi(t)A\phi(t')}} = -2 \frac{\partial}{\partial A_{22}} \log \int DxD\hat{x}e^{-\frac{1}{2}\phi A\phi} \]
\[ = -2 \frac{\partial}{\partial A_{22}} \log(\det A)^{-\frac{1}{2}} = \frac{\partial}{\partial A_{22}} \log(\det A) = (A_{22}^{-1}) \] (8.45)
\(^1\)We distinguish the advanced and retarded propagators. The former is zero if \( t \) is to the past of \( t' \), the latter is zero if \( t \) is to the future of \( t' \).
8.4. A SIMPLE EXAMPLE: THE HARMONIC OSCILLATOR

and similarly one can derive (8.41, 8.42, 8.43).

In order to obtain $C, G_A$ and $R$, there are two ways.

First way

Go directly to Fourier space: $-i\partial_t \rightarrow \omega$.

\[
A(\omega) = \begin{pmatrix} D(\omega) & \omega - im \\ -\omega - im & 0 \end{pmatrix} \rightarrow A^{-1}(\omega) = \frac{1}{\omega^2 + m^2} \begin{pmatrix} 0 & -\omega + im \\ \omega + im & D(\omega) \end{pmatrix} = \left( \begin{array}{c} 0 \\ \frac{1}{\omega^2 + m^2} \end{array} \right) (8.46)
\]

Second way

In matrix notation

\[
I = AA^{-1} = \begin{pmatrix} D & -i\partial_t - im \\ i\partial_t - im & 0 \end{pmatrix} \begin{pmatrix} G_A & G_R \\ 0 & C \end{pmatrix} = \begin{pmatrix} -(i\partial_t + m)G_A & DG_R - (i\partial_t + im)C \\ (i\partial_t - im)G_R & (i\partial_t - im)G_R \end{pmatrix} = \left( \begin{array}{c} \delta(t - t') & 0 \\ 0 & \delta(t - t') \end{array} \right) (8.47)
\]

Going to Fourier space and equating the elements of the matrices we have

\[
-(i\partial_t + m)G_A = \delta(t - t') \rightarrow (\omega - im)G_A(\omega) = 1 \quad (8.48)
\]
\[
(i\partial_t - im)G_R = \delta(t - t') \rightarrow -(\omega + im)G_R(\omega) = 1 \quad (8.49)
\]

the first two equations of (8.44) follow. The third one is easily verified by observing that $C(\omega) = D(\omega)/(\omega^2 + m^2)$ solves

\[
D(t, t')G_R - (i\partial_t + im)C(t, t') = 0 \rightarrow -\frac{D(\omega)}{\omega + im} + (\omega - im)\frac{D(\omega)}{\omega^2 + m^2} = 0 \quad (8.50)
\]

Going back to time domain we have

\[
G_A(t - t') = \int d\omega e^{i\omega(t-t')}G_A(\omega) = \int d\omega e^{i\omega(t-t')} \frac{1}{\omega - im} = e^{m(t-t')}\theta(t - t') \quad (8.51)
\]

where in the last step we performed an integration by residue, noting that the integrand has a singularity at $\omega = im$ and closing the integration path in the positive semiplane.

For the correlation function we have

\[
C(t - t') = \int d\omega e^{i\omega(t-t')} \frac{D(\omega)}{\omega^2 + m^2} = i \lim_{\omega \rightarrow \pm im} (\omega \mp im)e^{i\omega(t-t')} \frac{D(\omega)}{\omega^2 + m^2}
\]
\[
= \begin{cases} 
  e^{-m(t-t')} \frac{D(im)}{2m} & t > t' \\
  e^{m(t-t')} \frac{D(-im)}{-2m} & t' > t
\end{cases} \quad (8.52)
\]
Now we use $D(t, t') = 2T \delta(t - t')$, so $D(\omega) = 2T$, independent of $\omega$. So for all $t$

$$
C(t, t') = e^{-m|t-t'|} \frac{T}{m} \tag{8.53}
$$

We retrieve the fluctuation-dissipation theorem:

$$
\theta(t-t') \partial_t C = -T e^{-m|t-t'|} \theta(t-t') = -T G_A \tag{8.54}
$$

8.5 Quenched disorder

So far we have looked only at one-dimensional stochastic processes. The application of the thoery to multi-component processes is straightforward. We now consider a system of $N$ microscopic particles, which we label by Roman indices, $i = 1, ..., N$. The state of every microscopic element evolves in time according to the following Langevin equation

$$
\dot{\phi}_i(t) = f_0(\phi_i(t)) + f_i^J(\phi(t)) + \eta_i(t) \tag{8.55}
$$

where the term $f_i^J$ contains all the interactions between the different particles and particle $i$. The interactions are random and fixed during the life of the system, so that they are referred to as “quenched” disorder.

The noise $\eta$ does not depend on the quenched disorder and is again assumed Gaussian

$$
\langle \eta_i(t) \eta_j(t') \rangle = 2T \delta(t - t') \delta_{ij} \equiv D_0^i(t - t') \delta_{ij} \tag{8.56}
$$

We define

$$
\mathcal{L}_i^0(\phi_i(t)) = \partial_t \phi_i(t) - f_0(\phi_i(t)); \quad \mathcal{L}_J = -f_i^J(\phi(t)) \tag{8.57}
$$

The Langevin equation now reads

$$
\mathcal{L}_i^0 + \mathcal{L}_i^J = \eta_i(t) \tag{8.58}
$$

and following the same steps as before, the generating functional is given by

$$
Z[\psi] = \int \prod_i D\phi_i D\dot{\phi}_i \exp \left\{ \sum_i \int dt' i \psi_i(t) \phi_i(t) \right. \\
+ \sum_i \int dt \dot{i} \phi_i(t) \left[ \mathcal{L}_i^0 + \mathcal{L}_i^J \right] - \frac{1}{2} \int dt dt' \sum_i \dot{i} \phi_i(t) D_0^i(t-t') \dot{i} \phi_i(t') \left. \right\} \tag{8.59}
$$

In dynamics $Z[0]$ plays the same role as partition function in thermodynamics. This would suggest to average over the quenched disorder log $Z$ rather than $Z$ itself. However, since $Z[0] = 1$, $Z$ can be safely averaged over the quenched disorder and there is no need for replicas in dynamics. Time will play role of replicas, though. By averaging over the quenched disorder we decouple the sites but couple different times.

$$
\overline{Z[\psi]} = \int \prod_i D\phi_i D\dot{\phi}_i \exp \left\{ \sum_i \int dt' i \psi_i(t) \phi_i(t) + \sum_i \int dt \dot{i} \phi_i(t) \mathcal{L}_i^0 - \frac{1}{2} \sum_i \dot{i} \phi_i(t) D_0^i(t-t') \dot{i} \phi_i(t') \right\} \\
\times \exp \left( \sum_i \int dt \dot{i} \phi_i(t) \mathcal{L}_i^J \right) \tag{8.60}
$$
Now the disorder is only contained in the exponent and can be integrated out. This average will renormalize the coefficients of $\hat{\phi}^2$ and $\hat{\phi}$ giving rise to a new effective Langevin equation. Define

$$e^{\Delta[\phi, \hat{\phi}] \equiv e^{\sum_i \int dt \hat{\phi}_i(t) L_i}}$$ (8.61)

Once the average is done, it will be possible to isolate various terms in $\Delta$:

$$\Delta[\phi, \hat{\phi}] = -\frac{1}{2} \int dt \int dt' \sum_i \hat{\phi}_i(t) D_0(t-t') \hat{\phi}_i(t') + i \sum_i \int dt \hat{\phi}_i(t) L^1 + \cdots$$ (8.62)

$L_1$ renormalizes the disorder independent part of the Langevin equation and $D_1$ renormalizes the noise correlator

$$Z[\psi] = \prod_i \int D\phi_i D\hat{\phi}_i e^{-\frac{1}{2} \int dt \int dt' \hat{\phi}(t)(D_0 + D^1) \hat{\phi}(t')} + \int dt \int dt' \hat{\phi}(t)(L^0 + L^1) + \int dt \int dt' \hat{\phi}(t) \phi(t)$$ (8.63)

So we have transformed our original problem into the problem of solving the dynamics of a system evolving according to the effective Langevin equation

$$L^0 + L^1 = \xi(t)$$ (8.64)

with $\langle \xi \xi \rangle = D_0 + D^1$ (8.65)

In other words, the disorder is no longer present in our system, but the new noise $\xi$ is no longer $\delta$-correlated in time. The integration over the disorder has introduced a sort of memory in the dynamics of the system. This is a common phenomena in statistical physics: whenever starting from a Markovian stochastic process we integrate over some degree of freedom (disorder, fast variables, momenta, etc.) we end up with a new effective equation which is no longer Markovian and where modes which were previously uncoupled are now coupled.
Appendix A

Kramers-Kronig relations

Let us study before the consequences of causality. Causality of \( R(t) \) \((R(t) = 0 \forall t < 0)\) allows to define the Laplace transform

\[
\tilde{\chi}(z) = \int_0^\infty dt \, e^{zt} \, R(t)
\]

as an analytic function for any complex value of \( z = z_1 + iz_2 \), with \( z_2 > 0 \). Function \( \tilde{\chi}(\omega) \) is complex \((\tilde{\chi}(\omega) = \tilde{\chi}^*(-\omega))\) but causality enables us to obtain a relation between its real and imaginary parts. We do this by introducing the following trick. Integrate the function

\[
f(z) = \frac{\tilde{\chi}(z)}{z - \omega}
\]

over a contour \( C \) so that the contour does not enclose any pole of \( f(z) \). Since \( \tilde{\chi}(z) \to 0 \) as \( z_2 \to \infty \) there will be no contribution from the semicircle at infinity. Thus

\[
\int \frac{\tilde{\chi}(z)}{z - \omega} = \int_{-\infty}^{\omega-r} d\omega' \, \frac{\tilde{\chi}(\omega)}{\omega' - \omega} + \int_0^{\infty} d\omega' \, \frac{\tilde{\chi}(\omega)}{\omega' - \omega} + i\pi \lim_{r \to 0} \int_0^\pi d\phi \, \tilde{\chi}(\omega + re^{i\phi}) = 0
\]

It is useful to introduce the Cauchy principal part

\[
P \int_{-\infty}^{\infty} d\omega' \, \frac{\tilde{\chi}(\omega)}{\omega' - \omega} = \lim_{r \to 0} \left[ \int_{-\infty}^{\omega-r} d\omega' \, \frac{\tilde{\chi}(\omega)}{\omega' - \omega} + \int_{\omega+r}^{\infty} d\omega' \, \frac{\tilde{\chi}(\omega)}{\omega' - \omega} \right] = i\pi \tilde{\chi}(\omega)
\]

or

\[
\tilde{\chi}(\omega) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} d\omega' \, \frac{\tilde{\chi}(\omega')}{\omega' - \omega}
\]

This equation is a consequence of causality and allows us to relate the real part \( \tilde{\chi}'(\omega) \) and the imaginary part \( \tilde{\chi}''(\omega) \) of the response matrix:

\[
\tilde{\chi}'(\omega) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} d\omega' \, \frac{\tilde{\chi}''(\omega')}{\omega' - \omega}
\]

\[
\tilde{\chi}''(\omega) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} d\omega' \, \frac{\tilde{\chi}'(\omega')}{\omega' - \omega}
\]

\[127\]
These relations are called the Kramers-Kronig relations and allow to calculate the real part of \( \tilde{\chi}(\omega) \) from the imaginary part and vice versa. As we have seen, the imaginary part can be obtained from experiments.